UNIVERSIDAD CARLOS III DE MADRID working papers

Working Paper 04-17
Economics Series 04
February 2004

Departamento de Economía
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 91 624 98 75

THE RENTAL EQUIVALENCE APPROACH TO NONRENTAL HOUSING IN THE CONSUMER PRICE INDEX. EVIDENCE FROM SPAIN*

Raquel Arévalo[1] and Javier Ruiz-Castillo[2]

Abstract

This paper presents new evidence from Spain that challenges the usual objections to the possibility of applying the rental equivalent approach to determine the weight that non-rental housing services should have in the CPI. Data from the EPFs (*Encuestas de Presupuestos Familiares*) for 1980-81 and 1990-91 permit a satisfactory explanation of market rents in terms of an index of housing quality, two geographical variables and the year of occupancy. These regression results provide a way to impute a rental value to non-rental housing units that takes into account the possible selection bias induced by systematic differences in housing characteristics between the market rental sector and the non-rental stock. On average, such hedonic values are not that different from the self-imputations provided in the EPFs by the occupants of such dwellings. Therefore, the consequences for inflation of using either of the two alternatives to assess the importance of non-rental housing in the CPI system are small. Instead, if non-rental housing services are dropped from the CPI, then it is estimated that the bias in the measurement of inflation during the 1995-2000 period would be 0.35% per year. The lesson is that, given the alternatives, eliminating non-rental housing services from the CPI -as is done at present in Spain and several other European countries- is an unnecessarily crude form of dealing with a difficult problem

[1] Universidad de Vigo
[2] Universidad Carlos III

* This paper contains material from the first author's Ph. D dissertation presented in the *Universidad Complutense* of Madrid in June 2001. It is a pleasure to acknowledge comments from Raquel Carrasco, Sergi Jimenez, and Gregorio Serrano, as well as from the members of the dissertation committee Olimpia Bover, Omar Licandro and Paloma Taltavull. However, any remaining shortcomings are our sole responsibility. Financial support from project BEC2000-0173 from the Spanish DGI is gratefully acknowledged.

I. INTRODUCTION

The treatment of housing in the Consumer Price Index (CPI) is one of the more vexing problems for theorists, official statisticians, central bankers and other users. Among European countries, for example, in Austria, Belgium, France, Greece, Italy, Luxembourg, Portugal, and Spain, owner-occupied housing services have been excluded from the CPI. This is also the situation of the Harmonized Index of Consumer Prices (HICP), the official indicator of inflation in the EU produced by Eurostat. The economic importance of housing services in household budgets makes this position a provisional solution.[1] This paper presents new evidence from Spain about the possibility of using the rental equivalent approach to determine the weight that non-rental housing services should have in the CPI. Two radically different versions of this approach lead to practically indistinguishable inflation rates for the 1985-2000 period. In comparison, dropping those services from the CPI leads to a considerable bias of about 0.35% per year in the measurement of inflation.

To introduce the issues involved, assume that a CPI of the usual type, that is, a fixed-weights, Laspeyres statistical price index for a whole country must be constructed. In a world in which all housing dwellings were rented in perfectly competitive markets, there would be no argument about how to treat this commodity in the CPI. The difficulty arises, of course, as soon as there is a large owner-occupied

[1] Recently, pressure to improve upon this situation has been also forthcoming from those who ask for asset prices, and especially housing stock prices, to have a role in the measurement of inflation. See the analysis and the references quoted in Goodhart (2001, p.F353), who forcefully concludes that "…the appropriate methodology for incorporating measures of housing price inflation into our overall statistics for inflation remains an urgent and important issue. It cannot be dismissed or ignored. It has to be addressed."

housing sector. In the absence of observable transactions between owners and users of housing services, it is not obvious at all how to determine the weight to be given to owner-occupied housing in the CPI, or exactly which prices should be monitored over time.

What is to be done? As is well known, the cost-of-living (COL) index is a price index that measures the change in consumption costs required to maintain a constant standard of living. From now on, accept that the theory of the COL index –or what is known as the economic approach to index number theory- provides the conceptual framework for the country's CPI.[2] This approach suggests that the relevant commodity to be included in the CPI is the flow of housing services provided by the owner-occupied stock. As Triplett (2001, p. F327) indicates, "The concept of consumption implies that the standard of living depends on the consumption of housing services, and not on the purchase of houses".[3]

Conceptually, to price the service flow provided by the non-rental housing stock one can follow what is known as the rental equivalent approach that consists of two parts.[4] First, to impute a value to the flow of housing services provided by the non-rental dwellings during the CPI's base period, there are two alternatives. One

---

[2] This is actually the case in the U.S., Netherlands, and Sweden (see United States Department of Labor 1997, Balk 1994, and Dalen 1999). The advantage of the COL approach is that it provides guidance based on consumption theory in practical issues like this. For in-depth discussions of the superiority of this position *versus* the 'not COL' approach advocated by Hill (1997) and Turvey (1999), see Triplett (2001) and Diewert (2000). For alternatives to the COL approach in the housing sector, such as the net acquisition and payment approaches, see the International Labour Organization (ILO) manual (Turvey *et al*. 1989), and Turvey (2000).

[3] As a matter of fact, the same problem arises with dwellings facilitated to the occupant as wages in kind or as a result of a public or a private transfer. Thus, in the sequel we will always refer to the entire non-rental housing stock.

[4] As pointed out in Diewert (2000), the rental equivalent approach can be traced back to Marshall (1897, p. 594) at least, and it is the approach taken for owner-occupied housing by the Bureau of Labor Statistics

could simply ask occupants (or experts) the rent they think that that the dwellings in question could carry in the rental market. This is what is done in Spain in the household budget surveys, known as *Encuestas de Presupuestos Familiares* (EPF), gathered by the *Instituto Nacional de Estadística* (INE) with the main purpose of estimating the official weights of the Spanish CPI. Alternatively, from information collected on housing characteristics for the whole stock and observed rents in the market rental sector, rental values for non-rental housing in the base period can be estimated using hedonic regression methods. Second, changes in non-rental housing costs can be conceivably estimated by changes in rents for housing of similar characteristics in the market rental sector.[5]

The appropriateness of the COL approach to price index practice has been repeatedly questioned by some statistical agencies.[6] The usual objection has usually taken two complementary forms. It is often said that rent controls or public subsidized rents interfere with the workings of the rental housing sector, and/or that the characteristics of the dwellings of a very thin market rental sector are not representative of those of a very large non-rental housing stock. Spain is a case in point. The market rental sector includes all privately owned dwellings rented after a 1964 law that liberalized first contracts on vacant units, and allowed the introduction of rent actualization clauses in contracts that retained an automatic renewal clause at

---

in the construction of the CPI in the U.S. since 1983 (see Gillingham and Lane, 1982), and by most countries in the world in the system of National Accounts (see Eurostat *et al.*, 1993).

[5] Within the economic approach, an alternative way of estimating the opportunity costs of non-rental housing is the user cost approach. See, for instance, Smith *et al.* (1988), Diewert (2000) and Triplett (2001), as well as note 35.

[6] According to Triplett (2001, p.327), "Beyond the rhetoric, the issue that drives much statistical uneasiness over the concept of the COL is the treatment of owner-occupied housing…It is perhaps an oversimplification to say that empirical problems in estimating the flow of services for owner-occupied

the tenant's discretion. However, part of the housing stock is still under the influence of compulsory renewal clauses and rent controls established during the 1920-1964 period. There is also a publicly subsidized rental housing sector. The vast majority of the housing stock is owner-occupied, and the percentage represented by the market rental housing sector during the last 25 years is only between 5-15% of the total stock. Nevertheless, the self-imputed rental values obtained in the 1973-74 and 1980-81 EPFs were used to determine the weight to be given to non-rental housing in the CPI system based on 1976 and 1983, respectively.[7] However, possibly because the prices of the housing stock had been through an upward cycle since Spain joined the European Union in 1986, the self-imputed values collected in the 1990-91 EPF were thought to be too high, and the non-rental sector was eliminated from the CPI system based on 1992.[8]

Therefore, the scene is set for the experiment conducted in this paper. In the first place, the hedonic approach to the task of imputing rental values to non-rental housing is applied to the information in the 1980-81 and 1990-91 EPFs. After a comparison with the traditional one, the hedonic procedure advocated in this paper consists of three steps. First, an index of housing quality is constructed by applying Multiple Correspondence Analysis to a set of physical attributes, or structural characteristics for the entire housing stock. Second, market rents are explained in terms of the housing quality index and geographical variables, controlling for the

---

housing have induced rejection of the COL index framework, but there is nevertheless considerable truth in the oversimplification".

[7] Self-imputations by occupants and expert judgments are also currently used in the U.S. and the Netherlands, respectively, to determine the official weight assigned to owner-occupied housing in the CPI.

[8] See Castro (1992), the spokesman for the INE at the time.

inverse relationship between a dwelling's observed rent and the number of years that it has been occupied. Third, after controlling for the possible selection bias induced by differences in housing characteristics between the market rental sector and the non-rental segment of the housing stock, these regression results are used to impute a rental value to each non-rental housing unit.

In the second place, to assess the plausibility of the two available versions of the rental equivalence approach favored in this paper, the hedonic rental values ("objectively") imputed to non-rental housing by the statistical procedures just described are compared with the self-imputations ("subjectively") suggested by the occupants.[9] Finally, the official inflation since August 1985 to December 2001 (according to the CPI systems based on 1983 and 1992) is compared with the inflation that would have obtained if the CPI weights for non-rental housing in a 57-dimensional commodity space had been constructed using our hedonic estimates.

The main empirical findings are the following four. (i) The explanation of rents in the post-64 market rental sector is quite satisfactory. After correcting for outliers, 58% of the variance in 1980-81 and 59% in 1990-91 is explained in terms of the housing quality index, two geographical variables and the year of occupancy. It should be emphasized that, possibly because of the slow introduction of rent actualization clauses, sitting tenants with long contracts enjoy very large discounts relative to the rents paid in new contracts signed during the survey period for similar housing units. (ii) Dwellings in the market rental and the non-rental housing sectors turn out to have rather similar characteristics. Therefore, the estimation of a final model for market rents taking into account the interdependence between the choices of tenure mode and housing characteristics is not unduly affected by selection bias. (iii) There is broad agreement between the hedonic rental values and the self-evaluations for non-rental housing services. On average, hedonic values in 1980-81 are 10% higher, and in 1990-91 15% lower, than self-imputed ones. (iv) This leads to the most remarkable result of the paper: using either of the two alternatives to assess the importance of non-rental housing in the CPI system has very small consequences for inflation. Instead, if the rental equivalence approach is abandoned and non-rental housing services are dropped from the CPI, the weight of the "non-rental housing" commodity goes down by about 10 percentage points. As a consequence, inflation would be 0.33% per year higher from August 1985 to December 1992, and 0.38% per year lower from January 1993 to December 2000.[10]

The rest of the paper consists of 5 Sections and an Appendix. Section II presents the minimum information on institutional background and data sources that are necessary to understand the task at hand. Section III contrasts two alternative strategies to the question of explaining market rents within the hedonic approach: the traditional one that, together with geographic characteristics and the year of occupancy, uses a set of physical attributes as regressors; and a second view in which, as explained in the Appendix, a housing quality index is first constructed for the entire housing stock applying Multiple Correspondence Analysis to the physical attributes. Section IV obtains a final set of regression results that takes into account

---

[9] This is what is done in Francois (1989) who used the traditional hedonic approach with a sample of new tenants only. For his results, see below note 30.

[10] For comparison purposes, recall that when all sources of bias are considered, according to the Boskin (1996) Commission the official CPI in the U.S. suffers from and estimated upward bias of 1.1% per year.

the possible interdependence between the choices of tenure mode and physical attributes. Section V compares the hedonic values imputed to non-rental housing with the help of these regression results with the self-imputed values provided by their occupants. Section VI concludes.

II. INSTITUTIONAL BACKGROUND AND DATA SOURCES

To understand the problem addressed in this paper, it is essential to classify the housing stock according to both tenure mode and the legal status determined by government interventions in the housing sector. As far as tenure mode is concerned, dwellings occupied by their owners and rented dwellings constitute the owner-occupied and the rental sector, respectively. In the third mode, occupants do not own or rent the dwellings, but use them either as wages in kind, or as a transfer from a public organism, a private institution or an individual person. This will be referred to as the "Other" mode.

As indicated in the Introduction, the data sources for this paper are the 1980-81 and 1990-91 EPFs collected from April 1980 to March 1981, and from April 1990 to March 1991, respectively. These two household budget surveys consist of 23,971 and 21,155 observations, representative of the household population occupying non-institutional, or residential housing in all of Spain. This population is around 10 million and 11.3 million households in 1980-81 and 1990-91, respectively. The INE provides a set of blowing up factors to convert sample statistics into population statistics.[11] The reliability of the housing information provided by the 1980-81 and 1990-91 EPFs has been assessed in Arévalo (2001) by a comparison, whenever possible, with the corresponding information obtained from a 25% sample of the 1981 and 1991 Housing Censuses.[12]

As can be seen in Table 1, both EPFs heavily underestimate the amount of secondary housing recorded in the Censuses.[13] However, the coverage of permanently occupied residential housing by the EPFs is very satisfactory: approximately 96% both in 1980-81 relative to the 1981 Census, and in 1990-91 relative to the 1991 Census. The distribution by tenure mode according to the two data sources is very similar in the two periods. Taking also into account that the housing information for secondary housing in the EPFs is not as good as for permanently occupied residential housing, only the latter will be studied in the sequel.[14]

Table 1 around here

In Spain, as in other countries, government intervention in the housing sector takes many forms. For our purposes, it will suffice to describe the stylized features of two major policies. In the first place, the public sector has entirely financed some

---

[11] All the information used in this paper has been made accessible in http://www.eco.uc3m.es/epf80-81.html and http://www.eco.uc3m.es/epf90-91.html. For sampling methods and a full description of the data, see INE (1983, 1992a).

[12] These Censuses investigate the entire housing stock in March 1 1981 and 1991, respectively. See INE (1982, 1992b).

[13] As pointed out in Arévalo (2001, Chapter 1), this is partly due to the fact that the Census is dated one year after the beginning of the corresponding EPF. Moreover, the EPFs only investigate permanently or temporally occupied housing, while the Censuses also cover all unoccupied housing; respondents to the Census may very well classify part of the latter as temporally occupied , or secondary housing.

[14] As far as other comparable housing characteristics, relative to the Censuses both EPFs underestimate the proportion of residential housing in rural municipalities with less than 2,000 inhabitants. The main puzzle has to do with housing age: the proportion of dwellings more than 30 years old is overestimated in the 1980-81 EPF, but underestimated in the 1990-91 EPF.

housing construction but, starting from the 1950s, most government intervention has operated through a variety of programs that provide incentives to housing construction by means of moderate direct subsidies, low interest rates and/or fiscal credits to the producers. Complex public regulations, the details of which are unnecessary to enter into here, mandate that housing built under these policies are rented or sold below market prices. In addition, buyers of public housing may have access to more favorable financial conditions than if they had acquired their home in the unsubsidized private sector. As a result of this type of government intervention, housing units in the owner-occupied or the rental sectors will be classified as public or private housing.

Within the owner-occupied sector, all private housing units integrate what might be called the market sector. However, due to the public policy known as rent control, in the rental sector this is not the case. Together with the compulsory renewal clause on all leases, which dates from 1920, a 1946 law systematically froze all rentals at the level reached at the time of the first contract, following up on transitory regulations in the same vein already in effect in previous years. A 1955 law made a dramatic policy change, allowing for an almost unrestricted bargaining on all contracts made after May 12, 1956. Further legislation, enforced since July 1, 1964, sanctioned and extended that policy change. Lease renewals were still compulsory, but rents in new contracts are allowed to be determined by market forces; owners and renters were also allowed to include rent revision clauses subject only to annual ceilings set by the government.[15] However, for all housing rented before 1964, the power of rent revision remained with the government. Although this power has been exercised on several occasions since that date, only moderate increases have been permitted. Therefore, a distinction must be made between private rental housing occupied before or after the crucial date of 1964, giving rise to pre-64 or rent controlled housing, and post-64 or market rental housing, respectively. Within the period covered in this paper, which extends until 1990-91, legislation enforced since April 1986 did away with compulsory renewal clauses and completely liberalized the rental sector.[16]

Information on the legal status of residential housing is very hard to come by. Current occupants may not know or remember whether their dwellings were originally constructed under a public program. Consequently, Censuses do not even attempt to distinguish between private and public housing. Thus, this statistical source does not distinguish either between pre-64 and post-64 private rental housing. Fortunately, as can be seen in Table 2, the EPFs provide some partial but valuable information about the legal status of housing units in the rental and the owner-occupied sectors.

Table 2 around here

In recent decades, a revolution in tenure modes has taken place in Spain. The strict rent control up to 1964, the impediments remaining after the liberalization in that date until very recently (compulsory renewal clause and government limits to

[15] The post-64 situation in Spain belongs to what has been termed second-generation rent control in Arnott (1995) or tenancy rent control in Basu and Emerson (2000).

[16] A new law passed in 1995 reinstates some protection to tenants. Annual rent renewal clauses are authorized, but tenants can force all contracts to last at least five years.

annual rent increases), and the uncertainties created by the possibility of further legislative change, have turned the construction industry toward the owner-occupied sector. Moreover, a public policy of tax benefits to private suppliers has deliberately not worked in the direction of redressing the balance in favor of rental housing. Finally, since its inception in 1977 in the midst of the democratic transition started in 1975, the personal income tax has provided large incentives for the taxpayer to channel savings towards housing investments. As a result of these factors, according to Census data the share of rental housing has declined from 52% in 1950, to 30% in 1970, 21% in 1980, and 15% in 2000.

Against this background, the problem addressed in Sections III and IV is how to use the information on housing characteristics and rents in the market rental sector (row 1 in Table 2, i.e., 2,181 and 1,061 observations in 1980-81 and 1990-91, respectively), to impute a rental value to all housing units in the non-rental housing sector for which the EPFs provide self-imputations made by their occupants (rows 5 to 7 and row III in Table 2, or 18,487 and 18,180 observations in 1980-81 and 1990-91, respectively).

III. EXPLAINING MARKET RENTS WITHIN THE HEDONIC APPROACH

This section consists of three parts. The first one introduces the statistical model associated with the traditional hedonic approach to the explanation of market rents in terms of physical attributes and location or geographical variables. As time goes by, some housing units become vacant while others remain occupied. It is shown that, as long as rent increases incorporated in new contracts for vacant units are not exactly matched through rent renewal clauses in old contracts for dwellings already occupied, the year of occupancy would have to be included as an explanatory variable of the survey year market rents.

The second part presents the empirical models for the 1980-81 and 1990-91 Spanish samples. With regard to the first group of variables, namely, the physical attributes, two alternatives are contrasted in 1980-81: (a) the standard hedonic model in which all physical attributes enter linearly in the regression but no interaction between them and the year of occupancy is considered, and (b) a model in which the housing quality index constructed in the Appendix is substituted for the set of physical attributes and interactions between the index and the year of occupancy are included. Given its superiority on statistical grounds, only the second approach is applied in the 1990-91 case. The estimation process is completed in all cases with a diagnostic analysis, based on Peña and Yohai (1995), to detect potentially influential observations, to actually measure their influence in the regression results, and to assess their statistical significance as outliers. The final part of the section provides an economic interpretation of the empirical results.

III.1 The Statistical Model Associated With the Hedonic Approach

In the hedonic approach to the study of heterogeneous, differentiated commodities, unobservable quality differences between two product varieties are assumed to be well approximated by a set of observable physical attributes. This provided the original rationale to explain product prices in terms of product characteristics. Rosen (1974) establishes the microeconomic foundations of the approach by means of a perfect competitive model in which the price of an indivisible, differentiated product is jointly determined by the interaction of supply and demand of the product's attributes. Although the underlying demand and supply behavioral relations cannot be identified from the knowledge of product prices and product characteristics, partial derivatives with respect to each characteristic in a hedonic regression can be interpreted as the implicit marginal equilibrium price of the attribute in question.[17]

Let A_t be the annual rent in year t of those dwellings in the market rental sector whose first contract is made in that same year. Assume that the model for A_t is

$$A_t = F(t, \chi_t, \xi_t),$$

where $\chi_t = [\chi_{1t}, \chi_{2t}]$ is the set of physical (χ_{1t}) and geographical (χ_{2t}) attributes of housing units first rented in year t, and ξ_t is a white noise random term, normally distributed with $E[\xi_t] = 0$, $Var[\xi_t] = \sigma^2_\xi$, and $Cov[\xi_t,\xi_{t'}] = 0$ for all t and t'. Consider

---

[17] See Quigley (1979) for a survey of the early literature, and Sheppard (1999) for a recent one. Triplett (1990) contains a guide to the approach, as well as an illuminating account of why statistical offices have resisted its use for 30 years. Currently, after their success in the U.S. in the analysis of quality

the same housing distribution in the current period, χ_T, and, for simplicity, assume away all ageing effects. As long as $\chi_t = \chi_T$, $E[A_T / \chi_T] = E[A_t / \chi_t]$ only if there is no inflation in the housing sector. Assuming for illustrative purposes a constant rate π of housing inflation, we have

$$E[A_T / \chi_T] = E[A_t / \chi_t]\ (1 + \pi). \tag{1}$$

Let Λ_t be the subset of size N_t of those dwellings first occupied in year t which remain currently occupied in year T > t, and denote by a_t the vector of current rents in T. Assume that

$$a_t = f(t, X_t, \varepsilon_t), \tag{2}$$

where $X_t = [X_{1t}, X_{2t}]$ with $X_{1t} \in \chi_{1t}$ and $X_{2t} \in \chi_{2t}$ are the characteristics of the subset Λ_t, and ε_t is a white noise random term, normally distributed with $E\ [\varepsilon_t] = 0$, $Var\ [\varepsilon_t] = \sigma^2_\varepsilon$, and $Cov\ (\varepsilon_t, \varepsilon_{t'}) = 0$ for all t and t'. Notice that $E[a_t] = E[A_t / \chi_t = X_t]$ only if there is a rent freeze. In practice, A_t grows in time as a consequence of rent renewal clauses. Denoting by Δ the mean rate of rent increase due to renewal clauses from t to T, we have

$$E[a_t] = E[A_t / \chi_t = X_t]\ (1 + \Delta). \tag{3}$$

Comparing (1) and (3) it is easy to see that, on average, it would be indifferent to rent the housing stock with characteristics X_t in the current period T or in year t, if and

---

change in the context of the CPI, hedonic methods are at least widely discussed in statistical offices around the world.

only if renewal clauses exactly capture the impact of inflation. In this case, the mean impact of the year of occupancy on a_t would be zero. Otherwise, the year of occupancy should be an explanatory variable of the rent actually paid in year T of housing occupied in year t, as indicated in equation (2).

There are good reasons to expect a negative mean impact of the year of occupancy on present rents, meaning that owners would be giving up a discount to sitting tenants renewing their rents, relative to the rents charged in new contracts for housing of the same characteristics. In the first place, there are theoretical models yielding what is known as *tenure discounts*.[18] Landlords know that tenants will incur moving costs if they leave the unit. However, several factors work in the opposite direction. First, landlords also incur costs when a tenant moves out of a unit, including the cost of reconditioning, the cost of marketing a vacancy, and the rental income forgone during the vacancy. Second, landlords may want to retain current tenants that have shown during some period to be “good” tenants, minimizing wear and tear, avoiding trouble with neighbors, etc. Third, there could be a tenant’s decreasing willingness to pay with the passage of time. Thus, tenure discounts may appear as an equilibrium phenomenon in the game played by tenants and landlords. In the second place, it should be remembered that, in the Spanish case, new contracts signed between 1964 and 1986 have an automatic clause compelling landlords not to evict tenants up to two generations but in a very restricted set of circumstances. This shifts considerable bargaining power towards tenants in the rent renewal process, leading presumably to large tenant discounts (see Börsch-Supan 1986, Nagy 1997,

---

[18] See, *inter alia*, Guasch and Marshall (1987) and Hubert (1995).

and Basu and Emerson 2000). Moreover, rent increases for sitting tenants had to comply with an annual governmental ceiling, typically linked to the previous year's housing inflation rate.

Therefore, as Hoffman and Kurz (20002) conclude, both the peculiarities of housing markets and the regulations may result in tenancy discounts, which may cause a kind of lock-in effect with local non-substitution, since old contracts are not available to potential new tenants, and new contracts may not be attractive to sitting tenants, even if the old unit does not suit their needs anymore. At any rate, there is ample empirical evidence on tenure discounts in several countries under different legal arrangements.[19]

III.2. Empirical Results

There are three sets of empirical results corresponding to (i) a traditional hedonic model for 1980-81 that explains market rents in terms physical attributes, geographic characteristics and year of occupancy; (ii) a hedonic model for that sample in which the physical attributes are replaced by the housing quality index constructed in the Appendix, and (iii) a hedonic model of the latter type for 1990-91.

The Standard Hedonic Model Without Interactions Between Physical Attributes and the Year of Occupancy, 1980-81

Assume that in 1980-81 there is a sample consisting of n_t observations, with $n_t < N_t$, of housing units first rented at time t = 1965,…, 1980-81, so that the sample size is

[19] For the U.S., see for instance, Lowry (1981), Goodman and Kawai (1985) and Clark and Heskin (1982). For Germany, see Börsch-Supan (1986), Schlitch (1983) and Hoffman Kurz (2002). For Spain, see Peña and Ruiz-Castillo (1984).

$n = \Sigma_t n_t < N = \Sigma_t N_t$. Denote by $\mathbf{a} = \cup_t\{a_t\}$ the set of rents actually paid in 1980-81, $\mathbf{X}_1 = \cup_t\{X_{1t}\}$ the set of physical housing attributes, $\mathbf{X}_2 = \cup_t\{X_{2t}\}$ the set of geographic characteristics, $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2]$ the set of housing characteristics of both types, and $\mathbf{X}_3$ = {Ocup65, Ocup66,..., Ocup80-81} an index set of years of occupancy where, for each t, Ocupt = 1 if the housing unit was first rented in year t and Ocupt = 0 otherwise. Each housing observation, indexed by i = 1,..., n, can be described by

$$(a_i, X_i, X_{3i}) \in \mathbf{a} \times \mathbf{X} \times \mathbf{X}_3.$$

Assume for the time being that the impact of physical and geographic characteristics on rents is independent of the year the housing unit was first rented. That is, in terms of equation (2), assume that $\partial f/\partial X_t = \partial f/\partial X_{t'}$ for all $t \neq t'$. Under this simplifying assumption, rather than working with 16 separate models $a_t = f(t, X_t, \varepsilon_t)$, t = 1965,..., 1980-81, it is possible to work with a single one:

$$\mathbf{a} = g(\mathbf{X}, \mathbf{X}_3, \varepsilon), \tag{4}$$

where ε is a normally distributed random term with $E[\varepsilon] = 0$ and $Var[\varepsilon] = \sigma_\varepsilon^2$.

According to the 1980-81 EPF, the number of housing units in the post-64 or market rental sector is 2,181 (see row 1 in Table 2). However, in 18 cases there is no rent information, while in 21 cases there is no information on the building age. After dropping these 39 observations, the actual sample size becomes 2,142, representative of 867,627 population units. Descriptive statistics for all variables are presented in Table 3.

Table 3 around here

There are 7 structural characteristics in the set $\mathbf{X}_1$, including 5 discrete variables whose categories are always ordered so that the more desirable ones are assigned higher numbers. There are 3 categories of hygienic services (after some aggregation of an original list of 7 categories); 4 categories describing water facilities; 3 categories describing heating facilities, and 2 dummy variables indicating the presence of garage and telephone, respectively. The two continuous variables are housing area, measured in square meters, and building age, measured by the number of years between 1980 and the construction year. In Table 3, both variables have been discretized into 4 and 5 categories, respectively. There are two geographical variables in the set $\mathbf{X}_2$: municipal size, measured by the number of inhabitants, and the province where the dwelling is located. The 52 Spanish provinces have been classified into 4 groups, described in Table 3, having a similar mean housing price per square meter in 2002 (see *Ministerio de Fomento* 2002: http/www.mfom.estadisticas). Table 3 also includes the 16 dummy variables describing the distribution by year of first occupancy. For each variable, the population frequency and the mean monthly rent paid in 1980-81 (in euros) within each category is provided. In all cases, the more desirable the category, the larger the mean rent paid during the survey year is.

The inspection of the sample rent distribution conducted in Arévalo (2001), as well as the residuals of some preliminary linear specification of model (5), led to the following semi-logarithmic functional form with parameters (α, β, γ)[20]:

$$\ln \mathbf{a} = \alpha + \mathbf{X}\,\mathbf{b} + \Sigma_t\, \gamma_t \text{ ocupt} + \varepsilon. \qquad (5)$$

Table 4 contains the results of the estimation by OLS with robust standard errors of 4 versions of this general specification. To begin with, only the effect of variables in $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2]$ is studied. After some experimentation with alternative discretizations of all variables, it turns out that the logarithm of housing size (Lnsize) and the building age (Age) should enter in continuous form, but the following categories in Table 3 are non-significant: Heat2, Mun5, Prov1, Prov2, and Prov3. Therefore, they are all eliminated from the regression so that they only affect the constant in Model I. The remaining 14 variables (except the presence of garage and telephone facilities) are significant and appear with the expected sign (see below for a discussion of the role of each of them in the final Model IV).

Table 4 around here

The next step is devoted to the effect of the year of first occupancy. Model II includes 15 dummy variables, Ocupt with t = 65,…, 79. This reduces the mean square error of Model I by 11%, and raises the R^2 from 0.361 up to 0.499. As far as the effect on the $\mathbf{X}_1$ and $\mathbf{X}_2$ variables, the coefficients' size are slightly reduced in absolute terms in all cases, except for the variable Phone that becomes significant. Except for

[20] There is a large literature on the appropriate choice of functional form for the hedonic price function (see the discussion in Sheppard 1999, for example), but the simple log-linear form generally performs well.

Ocup79, all other occupational dummies have a significant effect on rents with the expected sign, indicating the existence of sizable tenure discounts. In absolute terms, except for Ocup65 and Ocup70 the further into the past the year of occupancy, the larger the γ_t coefficient is, but the observed differences do not justify such a large disaggregation level. The residuals of Model I are positively related to the number of years of occupation, with a stronger relationship since 1974 (not shown here). The relationship in Table 3 between mean rent values paid in 1980-81 and years of occupancy also shows two discontinuities at 1974 and 1978. To simplify the way in which this important variable enters into the analysis, Arévalo (2001) distinguished its effect during the sub-periods 1965-1973, 1974-1977, and 1978-1980 by means of 6 variables: 3 dummy variables, each of which takes the value 1 in one of the sub-periods and 0 otherwise; and 3 continuous variables, each of which is equal to the number of years of occupancy during the relevant sub-period and zero outside of it.[21] After some experimentation (whose results are available on request), it was found that it is unnecessary to distinguish the third sub-period. Therefore, the best specification, shown in Model III, includes a single dummy variable Ocup6573 (so that the dummy eliminated from the regression equation is Ocup7480), and 2 continuous variables, Year6573 and Years7480. The goodness of fit of this model is similar to the one for Model II, but there are 12 fewer parameters to estimate. The regression coefficients of the X_1 and X_2 variables remain essentially constant.

[21] Thus, for example, a housing unit first rented in 1970 is characterized by a value of 1 in the dummy Ocup6573, a value of 10 in the continuous variable Years6573, and a value of 0 in each of the 2 remaining dummy variables, Ocup7477 and Ocup7880, and the 2 remaining continuous variables Years7477 and Years7880.

According to Peña and Yohai's (1995) method, there are 90 influential observations, or 4.2% of the sample, that can be considered outliers[22]. Model IV in Table 4 is the final model after deleting all outliers. The mean square error is reduced by 16.4%, while the R^2 rises up to 0.586. Except for Water4 and Mun3, the precision with which all $\mathbf{X}_1$ and $\mathbf{X}_2$ variables are estimated is improved upon and the presence of garage facilities becomes significant.

The Hedonic Model With a Quality Index, 1980-81

As explained in the Appendix, an (ordinal) housing quality index (Qindex) that summarizes the physical attributes has been constructed. Consider the regression equation where this index substitutes for the vector $\mathbf{X}_1$ of physical characteristics:

$$\ln \mathbf{a} = \alpha + \beta \qquad\qquad + \mathbf{X}_2\, \mathbf{b}_2 + \Sigma_t\, \gamma_t \text{ ocupt} + \nu. \qquad (6)$$

In this model, consistent with the hedonic approach, the β coefficient can be interpreted as the implicit marginal equilibrium price of housing quality.

There are several *a priori* reasons why this approach should be preferred to the traditional one.

(i) It permits overcoming the multicollinearity problem that may disturb the precise estimation of the physical attributes in vector $\mathbf{X}_1$ (recall, for example, that the category Heat2 had to be removed from the regression because it was non-significant).

---

[22] There are 40 observations with a t value between 3 and 5, 35 observations with a t between 5 and 7,

(ii) It permits estimating the impact of housing characteristics that, due to their infrequency, would have no explanatory power in the traditional approach. For example, in 1990-91 only 6 out of 1,026 housing units in the rental market sector have swimming pool facilities. In the second approach, this attribute may influence market rents through its effect on the housing quality index.

(iii) By construction, the quality index is uncorrelated with the remaining indicators in the Multiple Correspondence Analysis. In so far as these indicators are orthogonal to housing quality, they should have no explanatory power of a unit's market rent. Thus, using only Qindex in equation (6) filters possibly irrelevant information as far as explaining market rents is concerned.

(iv) To have a single variable, makes the study of interactions between the variables synthesized in the index and other explanatory variables considerably easier. As it will be seen below, this is the case in this context with respect to the year of occupancy.

Table 5 presents three regression models using Qindex in place of the 10 $\mathbf{X}_1$ variables that were found significant before (see Table 3 for some descriptive statistics of this new variable). Model A also includes the geographical variables in $\mathbf{X}_2$, as well as the best specification found before for the year of occupancy. The comparison of this model with Model 3 shows that, in spite of the reduction of the number of parameters, the goodness of fit is essentially preserved. The role of the geographical variables in both models is also very similar.

Table 5 around here

---

and 15 with a t greater than 7.

So far, in both approaches it has been assumed that the impact of the physical characteristics, or of Qindex, on rents is independent of the year the housing unit was first rented. Of course, in the traditional approach this assumption can be verified by interacting the variables in $\mathbf{X}_1$ with the year of occupancy. However, as anticipated in point (iv) above, the search for an interaction pattern should be much simpler with a single variable, Qindex, than with 10 of them. After some exploration in Arévalo (2001) –whose results are available upon request- the best specification is achieved by substituting two variables for Qindex: Qindex6573 = Qindex if $t \in [1965, 1973]$ and 0 otherwise, and Qindex7480 = Qindex if $t \in [1974, 1980]$ and 0 otherwise (see Table 3 for some descriptive statistics). The regression results, presented in Model B in Table 5, show that the positive relationship between market rent and housing quality is not constant over time, since it is distinctly stronger when the housing unit is rented after 1973. On the other hand, the coefficient for the dummy variable Ocup6573 remains negative but is not significant.

Interestingly enough, the application of Peña and Yohai’s (1995) procedure yields exactly the same 90 outliers already detected in the first approach. After deleting all the outliers, the results are in Model C of Table 5. The mean squared error is reduced by 16.2%, while the R^2 increases up to 0.578. Except for Mun3, the precision with which all variables are estimated increases and Ocup6573 becomes again significant.

The Hedonic Model With a Quality Index, 1990-91

Descriptive statistics for all variables in $\mathbf{X}_1$, $\mathbf{X}_2$, and $\mathbf{X}_3$ are presented in Table 6. As we have just seen, using only Qindex in 1980-81 as in equation (6), rather than the 10 variables in $\mathbf{X}_1$as in equation (5), considerably reduces the number of parameters to be estimated without damaging the regression's goodness of fit, and makes it much easier to model the interaction between housing quality and year of first occupancy. As can be seen in Table 6, besides the 8 housing characteristics included in 1980-81, in 1990-91 there is information on 10 more physical attributes. Therefore, the above advantages are expected to be even more important in this case. This justifies adopting the Qindex approach in 1990-91.

Table 6 around here

Model A in Table 7 uses the available observations in the market rental sector, namely, 1,061 observations (see row 1 in Table 2), less 19 without information on rent, and 7 without information on building age; that is, a total of 1,035 observations representative of 590,948 housing units at the population level. As before, the hardest issue is how to model the year of occupancy's effect. Preliminary explorations indicate different behavior during 3 different sub-periods. This leads to 3 continuous variables, Years6575, Years7682, and Years8390, as well as 3 dummy variables Ocup6575, Ocup7682, and Ocup8390.[23] Correspondingly, to capture the interaction between housing quality and year of occupancy, 3 variables Qindex6575, Qindex7682, and Qindex8390 were created. As shown in Model A in Table 7, the best specification joins the variables Years7682 and Years8390 into a single one, called

[23] Thus, for example a dwelling first rented in 1980 is characterized by a value of 10 in Years7682, a value of 1 in Ocup7682, and a value of 0 in the remaining variables.

Years7690. According to Peña and Yohai's (1995) method, there are 53 influential observations, or 5.1% of the sample, that can be considered outliers.[24] Model B in Table 6 is the final one after deleting all outliers. The mean square error is reduced by 15.6%, while the R^2 raises up to 0.592. Except for Years6575 and Qindex8390, which remain nevertheless highly significant, the precision with which all variables are estimated is improved upon and Years7690 and Ocup6575 become significant. On the other hand, the coefficients of the geographical variables in Model B display a similar pattern to the one shown in Table 5 for the 1980-81case.

Table 7 around here

III.3. Economic Interpretation

First, consider the traditional hedonic model IV in Table 4 for the 1980-81 sample. Individual coefficients provide a rich explanation of market rents. Starting with physical attributes, (a) to have less (or more) than one full bathroom leads to a 21.8% smaller (or to a 43.8% greater) estimated rent than in the reference situation. (b) Relative to having hot water from an individual system, centrally heated hot water increases rents by 33.4%, but to have no water at all or to have only cold water reduces rents by 40.0% and 20.9%, respectively. (c) Central heating increases rents by 25.6%, while the presence of (d) garage and (e) telephone facilities increases rents by 15.3% and 11.9%, respectively. (f) The dwelling size elasticity is 0.23, so that a 10% increase in size leads to an estimated 2.3% increase in rent. (g) The age of the building reduces rents by 0.3% per year.

---

[24] There are 33 observations with a t value between 3 and 5, 13 observations with a t between 5 and 7, and 7 with a t greater than 7.

When the 10 significant physical attributes are replaced by the housing quality index (see Model C in Table 5), it is observed that, as expected, greater quality implies larger market rents. This relationship between quality and rent is also preserved in 1990-91. However, the interaction between housing quality and year of occupancy displays opposite patterns in the two samples, a feature to which we will return below. At this juncture, notice that, as far as the geographical variables are concerned, the observed pattern is practically the same in all models: first, the greater the population of the municipality where the unit is located, the higher the estimated rent is; second, in 1980-81 rents are higher in Madrid, Barcelona and the other provinces where the housing stock has a mean price in 2002 higher than 1,700 euro per squared meter, while in 1990-91 rents are lower in those provinces where the housing stock has a mean price at that date lower than 860 euro per squared meter.

As has been noted above, the rent actualization process via rent renewal clauses for sitting tenants may proceed more slowly than rent increases in new contracts due to inflation. According to the hedonic model IV in Table 4, the annual discount generated by the difference between the sector’s inflation and the actualization clauses is 3.9% during 1965-73 and 12.7% during 1974-1980. The accumulated discounts on rents of dwellings of average quality occupied in different years, relative to units of the same quality first rented in 1980, are presented in column 1 in Table 8. They are very large indeed, ranging from 12% in a single year for units rented in 1979 to 70.1% for units first occupied in 1965. The accumulated discounts by year of occupancy according to model B in Table 5, where physical attributes are replaced by a single quality index, are presented in column 2 of Table 8. These refer to a unit of average quality with Qindex = 2.38 located in the area where housing stock prices are highest. The direct effects due to the year of occupancy are practically the same as in the traditional hedonic model, but model B includes an additional effect due to the interaction between housing quality and year of occupancy. As a result, the accumulated discounts are now slightly larger than before.[25]

Table 8 around here

To evaluate these results, it is necessary to turn our attention to the inflation that took place during this period in the housing rental sector. The Spanish INE measures the inflation rate of a sample of rental units that may include both private dwellings rented before 1964 and public housing dwellings, whose rents need not vary as those of private units rented after 1964.[26] The official inflation rate, reproduced in column 6 in Table 8, shows a structural change in 1973. The mean inter-annual inflation rate is 5.8% during 1965-72 and 12.5% during 1973-1980. Due to inflation, rents of dwellings of a given quality indexed at a value of 100 in 1965 would be 379.5 in 1980. If the occupants of those dwellings had enjoyed no rent actualization at all, they would have received an accumulated discount of 73.6% relative to the rent these units would have had in 1980, a figure very close to the ones in columns 1 and 2 in Table 8. The implication is that, relative to the inflation recorded in the entire rental sector, the estimated rent actualizations through renewal clauses from 1965 to 1980 have been negligible.

---

[25] Interestingly enough, the discount estimated in Peña and Ruiz-Castillo (1983) for a sample of housing units rented in the Madrid Metropolitan Area in 1975, is 8% per year. The accumulated discount for the 1965-1975 period is 56.56%, a larger figure than those of columns 1 and 2 in Table 8.

In 1990 things start to change (see Model B in Table 7). The direct effect of the year of occupancy has two structural changes in 1975 and 1982, but now the annual discounts decrease with time: they are 6.6% in 1965-75, and 2.7% in 1976-1990. Likewise, the coefficients of the quality index tend to decrease with time in the corresponding sub-periods. As a result, the accumulated discounts for dwellings of average quality (Qindex = 15.46) in the more expensive provinces first rented in years close to 1990 grow more slowly than what was observed before for years close to 1980 (see column 4 in Table 8). It would appear that, with the passage of time, an increasing number of new contracts do include rent renewal clauses. Nevertheless, the evidence of tenure discounts remains impressive. In particular, taking into account the official inflation rates, rents of dwellings indexed at a value of 100 in 1965 would be 85.94 in 1990. This means that, without any actualization at all, those units would have received an accumulated discount of 88.2% relative to the unit's rent in 1990, a figure still very close to the 83.5% estimated with our model.

IV. IMPUTING MONEY VALUES TO NON-RENTAL DWELLINGS

Given the regression results in Model C in Table 5 and Model B in Table 7 for 1980-81 and 1990-91, respectively, the next task is to impute in each survey year a monetary value to the flow of housing services provided by non-rental dwellings, namely, owner-occupied and transferred housing. However, if the choices of tenure mode and housing characteristics are not independent, the OLS estimation in the market rental sector might be inconsistent, as well as unbiased.

---

[26] For the difficulties and possible bias in the measurement of rent inflation in the rental housing

In order to deal with this problem, we begin by recognizing that the total sample of housing units, indexed by $i = 1,\ldots, N$, consists of two sets of observations: uncensored ones in the set $\{i = 1,\ldots, N | \; i \in \Lambda\}$, where Λ denotes the set of market rental units; and censored observations in $\{i = 1,\ldots, N | \; i \notin \Lambda\}$ for which market rents a_i are obviously not observed. Let z^* be the latent variable that determines the sample selection and it is only observable when $z^* \geq 0$, and let I be an indicator variable defined by

$$I_i = 1 \Leftrightarrow i \in \Lambda \Leftrightarrow z_i^* \geq 0, \text{ and } I_i = 0 \Leftrightarrow i \notin \Lambda \Leftrightarrow \mathbf{z}_i^* < 0.$$

In this situation, if $f(a, z^*)$ is the joint density function of the random variables a and z^* and $\Pr(z^* \geq 0)$ is the probability of belonging to the set Λ, then the observable conditional density demand function is

$$f(a, z^* | \; z^* \geq 0) = f(a, z^*) / \Pr(z^* \geq 0).$$

The implication is that the regression model (6) has to be reformulated introducing a selection mechanism for z^*. That is to say, we now have:

$$\ln a_i = \mathbf{Y}_i\mathbf{b} + \nu_i, \text{ if } i \in \Lambda, \tag{7}$$

and

$$z_i^* = \mathbf{Z}_i\mathbf{g} + u_i,$$

where, for all i, $\mathbf{Y}_i$ is a vector that includes the variables that have been shown to have a significant explanatory role in the previous section, namely a constant, the geographical variables, the year of occupancy variables, the housing quality index and its interactions with the year of occupancy; ν is normally distributed with $E[\nu_i]$ =

component of the CPI, see Randolph (1988), and Hoffman and Kurz (2002).

0 and $\sigma[\nu_i] = \sigma_\nu$; $\mathbf{Z}_i$ is the set of variables that permit classifying each housing unit according to its tenure mode, and u is a normally distributed variable with $E[u_i] = 0$ and $\sigma[u_i] = 1$. Therefore,

$$\Pr(i \in \Lambda) = \Pr(I_i = 1) = \Phi(\mathbf{Z}_i\mathbf{g}) \text{ and } \Pr(i \notin \Lambda) = \Pr(I_i = 0) = 1 - \Phi(\mathbf{Z}_i\mathbf{g}),$$

where Φ is the distribution function of a standard normal.

In order to get consistent and efficient estimates, the model is estimated by maximum likelihood methods, taking as initial values the results of the estimation of Heckman (1979) selection model in two stages. In this context, the log likelihood for each observation i is

$$l_i = \begin{cases} w_i \ln \Phi\big([\mathbf{Z}_i\mathbf{g} + (\ln a_i - \mathbf{Y}_i\mathbf{b})\rho/\sigma]/(1-\rho^2)^{1/2}\big) - (w_i/2)([\ln a_i - \mathbf{Y}_i\mathbf{b}]/\sigma)^2 - w_i \ln(2\pi)^{1/2}\sigma & \text{if } i \in \Lambda, \\ w_i \ln \Phi(-\mathbf{Z}_i\mathbf{g}) & \text{if } i \notin \Lambda, \end{cases}$$

where ρ is the correlation coefficient between the random terms ν and u, and w_i is the blowing up factor for each unit that permits to go from sample to population statistics. Given that the N observations are assumed to be independent, the function to be maximized is

$$L(\mathbf{b}, \mathbf{g}, \sigma \quad \rho | \; a, \mathbf{Y}, \mathbf{Z}) = \Sigma_i \, l_i. \tag{8}$$

Provided that the selection of the **Y** and the **Z** variables ensures the model's identification, the consistent estimation of the **b** parameters and the availability of the **Y** variables for all dwellings will allow the imputation of a monthly rent value to all units in the non-rental sector.

Together with the physical attributes that determine the housing quality index, the geographical variables and the year of occupancy, the set **Z** includes the migrant condition of the household, measured in the EPFs by means of a question asking whether the household had moved into the current municipality before or after 1975 or 1985 in the 1980-81 and the 1990-91 case, respectively. According to Table 2, there are 18,487 non-rental units in 1980-81 and 18,180 in 1990-91. However, 1,745 and 1,570 in these two periods lack information on building age, square meters or, above all, year of occupation.[27] Therefore, only 16,742 and 16,610 non-rental dwellings, representative of 7,044,514 and 8,782,425 million population units have been considered in 1980-81 and 1990-91, respectively. Descriptive statistics about all variables considered in the market rental and the non-rental sectors are in Table 9.

Table 9 around here

In 1980-81, non-rental units seem to have more hygienic services, water, garage, and telephone facilities; larger size and, in spite of the fact that all market rental dwellings have been occupied after 1964 but there is a sizable proportion of non-rental ones occupied before that date, the latter are less old than the rental market units. Relative to owner-occupied and transferred units, rented dwellings seem to be more prevalent in more populated municipalities and in the group of provinces with higher stock prices in 2000. The differences in all these dimensions,

[27] As many as 1,648 and 1,557 are occupied as wages in kind or because a private or a public transfer.

however, are not very large. In 1990-91 the pattern is exactly the same, including a greater proportion of detached units with garden facilities in the non-rental sector.

Not surprisingly, the proportion of migrant households living in rented accommodations is more that 4 times larger than those living in non-rental housing. Given that the migration status of the household does not appear as a determinant of housing rents, this is the variable that permits identification of model (7). The results of the estimation of model (8) in both years are presented in Table 10.

Table 10 around here

In 1980-81, but not in 1990-91, there is evidence that the sample selection bias must be corrected.[28] Therefore, when the coefficients of the variables explaining market rents in Table 10 are compared to those obtained in Section III (see Model B in Table 5 for 1980-81 and Model C in Table 7 for 1990-91), the differences are larger in 1980-81 than in 1990-91. Although the relative effect of all variables on market rents remain unchanged in both years, it is observed that the coefficients for the year of occupancy variables are slightly larger in absolute terms. This implies still larger discounts for sitting tenants in dwellings occupied before the sample year (compare columns 3 and 2 in Table 8 for 1980-81, as well as columns 5 and 4 in that same Table for 1990-91).

V. HEDONIC ESTIMATES vs. SELFIMPUTED VALUES

[28] As pointed out in Table 10, in 1980-81 the hypothesis of independence (H_0: p = 0) should be rejected for any confidence level below 99.25% (P-value = 0.0075). This is the case in 1990-91, but only for a confidence level below 82.6% (p-value = 0.1740).

As explained in the Introduction, for all non-rental units the EPFs record the monthly rent that an informer for the household occupying the unit thinks that his/her dwelling would command if it were rented in the market at the time of the survey. Answers to this question constitute what are called self-imputed rental values.

Given the parameter vector **b** estimated in the hedonic model (7), knowledge of the $\mathbf{Y}_i$ variables for a non-rental unit i will suffice to produce an imputed rental value dependent on the year of occupancy. However, to have a comparable imputation to the self-imputed value provided by the occupant, it is necessary to compute the rental value that the unit would command at the survey time, that is, making the variable year of occupancy equal to 1980 and 1990 in the two samples for 1980-81 and 1990-91, respectively. The resulting estimates will be referred to as hedonic rental values.[29]

There are 13 and 4 non-rental units missing self-imputed values in 1980-81 and 1990-91, respectively, that had to be eliminated from the comparison. The remaining 18,474 and 18,176 units with complete information, representing 7,772,078 and 9,602,498 population dwellings in 1980-81 and 1990-91, respectively, have been classified by the quartiles of the distribution of hedonic rental values. The frequency distribution of non-rental units by tenure mode and those quartiles are presented in the upper panel of Tables 11 and 12.

Tables 11 and 12 around here

[29] There are 20 non-rental units in 1980-81 and 233 in 1990-91 without information on building age, as well as 62 in 1990-91 without information on square meters. These have been assigned the mean value of

In 1980-81 more than 50% of non-rental units are part of the owner-occupied market sector, while slightly more than 25% are public housing. The remaining 20% is equally divided between owner-occupied housing of unknown legal status and other units occupied as wages in kind or as a result of a public or a private transfer. In 1990-91 the distribution by tenure mode is very similar, with a slight increase in the percentage of public housing and owner-occupied of unknown legal status and a corresponding decrease of the market sector and the fourth tenure mode. In both years, public housing is over-represented in the third and the fourth quartiles, while the other three tenure modes are over-represented in the second and, above all, in the first quartile.

The comparison between hedonic and self-imputed values takes two forms. In the first place, all units have been cross-classified by the quartiles of the two distributions. The middle panel of Tables 11 and 12 presents the diagonal terms of the corresponding contingency tables, that is, the percentage of units in each tenure mode that have been classified in the same quartile of the distributions of hedonic and self-imputed values. Naturally, the greater those percentages are, the closer the ranking of units is according to the two criteria in the corresponding sector. In addition, the middle panel of those Tables includes the Spearman coefficient, a statistic that provides a synthetic, scalar measure of the degree of association between the two rankings.

Generally, the agreement between the two classification criteria is somewhat better in 1980-81 than in 1990-91. For the non-rental stock as a whole, in these two

those variables in the tenure mode to which they belong. In this way, a housing quality index and an dates the Spearman coefficient is 0.66 and 0.57, respectively. In both years, the percentage of units equally classified is greater in the quartiles I and IV; apparently, it is harder to agree upon the relative ranking of dwellings of intermediate quality. By tenure mode, the maximum agreement is achieved in the larger sector, namely, the owner-occupied market sector. The minimum agreement takes place in the public sector. This is not surprising, given the high number and large variety of public housing policies developed in Spain since 1950. Taking into account the fundamental difference between the "objective" statistical procedure followed to assign hedonic values and the "subjective" nature of the self-imputations suggested by the occupants, it can be concluded that, on balance, the degree of agreement between the rankings of the non-rental housing units according to the two classification criteria is quite satisfactory in both years.

Beyond this ordinal analysis, the next question is about the differences in the mean values arrived at from the two routes. In the bottom panel of Tables 11 and 12 the differences in percentage terms between mean hedonic and self-imputed values by tenure mode are presented for the entire non-rental sector and the quartiles of the distribution of hedonic rental values.

In 1980-81, mean hedonic values are *10.6% greater* than mean self-imputed ones. The proximity of mean values in the first two quartiles is truly remarkable, but the discrepancies are somewhat larger in quartile IV. Clearly, in this occasion the main difficulty lies in assessing rental values for public housing dwellings of greater than average quality. In 1990-91, mean hedonic values are *15.3% lower* than self-

hedonic value for them have been computed.

imputed ones. This time the largest differences are in quartiles II and III. This seems to be due to the discrepancies found with respect to owner-occupied units of intermediate quality and private or of unknown legal status. As a matter of fact, differences in mean values in these sectors reach 25.2% and 18.6%, respectively.[30]

Unfortunately, there does not exist reliable statistical sources about housing prices at a national level prior to 1985.[31] However, there is no doubt about the existence of a continuous increase in housing prices starting around 1985, a year before Spain became a full member of the European Union, that lasts until 1991, shortly before the beginning of a downswing in the general business cycle for the Spanish economy. Evidence about housing prices for new transactions in Madrid presented in Bover (1993), speaks of a previous boom around 1979. Based on such tentative evidence, it would appear that the sign of the difference between hedonic and self-imputed values might be influenced by the phase of the housing cycle in which the self-imputations take place: a downswing in the case of the EPF collected from April 1980 to March 1981, and at the very end of an upswing for the EPF collected from April 1990 to March 1991.

Be that as it may, the important question is whether such differences are large or small for the purpose at hand. As explained in the Introduction, there are two ways to apply the rental equivalence approach to the determination of the weight that non-rental housing should receive in the CPI: using the evaluation by experts or

[30] Francois (1989) presents preliminary research within the BLS in the U.S. suggesting that owner estimates of implicit rents may be biased by as much 10% or 30% above actual values on average. However, he criticizes that research for underestimating spot rents, and points out that these results have not been confirmed by comparisons using data collector estimates. More importantly, using a sample of new tenants consisting of 3,706 observations in 27 different major metropolitan areas, this author shows that on average owner estimates are only lower than hedonic estimates by 1.4%.

occupants, or using the imputation obtained through hedonic methods. Therefore, in the context of this paper, the answer to the above question hinges on the implications for inflation of following either of the two versions of that approach.

Inter-annual inflation rates, Π_t, are computed using monthly CPI data according to the formula:

$$\Pi_t = (CPI_{m,t} - CPI_{m,t-1}) / CPI_{m,t-1},$$

where $CPI_{m,t}$ and $CPI_{m,t-1}$ are the value of the CPI in month m in years t and t – 1, respectively. In turn, the $CPI_{m,t}$ of the Laspeyres type with J commodities indexed by j = 1,..., J, is a weighted average of price relatives, $(p_{j,m,t} / p_{j,0})$, where $p_{j,m,t}$ is the price of commodity j in month m of year t, and $p_{j,0}$ is the price of that commodity during the base period 0. That is,

$$CPI_{m,t} = \Sigma_j W_j (p_{j,m,t} / p_{j,0}),$$

where W_j, j = 1,..., J, is the set of official CPI commodity weights. The Spanish INE uses the 1980-81 and 1990-91 EPFs to estimate the weights W_j of the CPI system based in 1983 and 1992, respectively: each commodity j is assigned a weight W_j equal to the ratio of the population expenditures on that commodity to the population total expenditures on all commodities. Since the INE publishes the price relatives $(p_{j,m,t} / p_{j,0})$ for certain commodity breakdowns, it is possible to use the EPFs to

[31] Bover (1993) contains a discussion of available statistical sources.

estimate a list of weights in order to reproduce the official inflation or to provide interesting alternatives.[32]

Among other classifications, in 1983 the INE uses a commodity breakdown with J = 57, where commodity 35 includes two main items: a “non-rental housing services” component, whose official weight is estimated using the self-imputed rental values declared in the 1980-81 EPF by these dwellings occupants; and a second component mainly consisting of “housing repair and maintenance”. Excluding secondary housing, for which as explained in Section II no quality index has been computed, our own estimates based on self-imputed values of the weights given to non-rental housing services and commodity 35 are 9.4% and 13.5%, respectively.[33] The estimates obtained with the hedonic values are 10.4% and 14.5%. Inter-annual inflation rates for these two alternatives, as well as the corresponding mean annual inflation for the entire period August 1985 to December 1992, are presented in the upper panel of Table 13. The differences are small (of the order of 0.03% per year). Therefore, whether we use self-imputations or hedonic values in assessing the importance of non-rental housing services in the CPI has a small impact on the measurement of inflation.

As an alternative, consider the possibility of abandoning the two versions of the rental equivalence approach. This means dropping altogether the non-rental housing services component from the CPI-as the Spanish INE did officially in the 1992 base. In this case, commodity 35’s weight becomes 4.6%. As can be seen in column 5 of Table 13, the mean inflation rate for the 1985-1992 period of the non-rental housing services

[32] For a thorough attempt in this direction in the context of a discussion of the Boskin Commission

component is almost 3 percentage points below the mean inflation rate for the economy as a whole. Therefore, as can be seen in column 3 in Table 13, inflation rates are now significantly lower than before. The differences with the rental equivalence approach are of a considerable order of magnitude. In particular, abandoning the rental equivalence approach represented by the hedonic alternative would have led to a bias in inflation of 0.33% per year during the 1985-1992 period (see column 4 in Table 13). For comparison purposes, recall the Boskin Comission's view that when all sources of bias are considered the official CPI in the U.S. suffers from an estimated upward bias of 1.1% per year.

Table 13 around here

In 1992, the INE uses a commodity breakdown with J = 57 where, apart from other minor changes relative to the 1983 base system, commodity 35 includes only housing repair and maintenance, local housing taxes and other items pertaining to non-rental housing, but excludes all imputations for non-rental housing services. In order to estimate the inflation rates that would be obtained if the rental equivalence approach were applied in the 1993-2000 period, there is no difficulty in creating an additional commodity 58 whose weight can be estimated using the hedonic or the self-imputed rental values for non-rental housing. The sum of the weights for commodities 35 and 58 in the hedonic and the self-imputed cases is 15.6% and 17.0%, respectively, while the weight of commodity 35 in our version of the official system is 4.8%.[34] In the absence of an official price index for commodity 58, the index for

findings in the Spanish case, see Ruiz-Castillo *et al.* (1999).

[33] The official values are 10.33% and 13.61%, respectively.

[34] The official weight itself is 5.30.% For a discussion of the reasons why our estimate differs from the official one, see chapters 2 and 3 of Ruiz-Castillo *et al.* (1999).

commodity 33, "rents paid in rental housing", has been used. The estimates of the 3 series of inflation rates, as well as the information about the evolution of commodity 33's prices, appear in the lower part of Table 13. Again, the differences between the two versions of the rental equivalence approach are small (about 0.05% per year). However, the downward bias incurred when the official system is used rather than the hedonic one is 0.38% per year –a considerable order of magnitude.

The conclusion is inescapable. The two versions of the rental equivalence approach lead to comparable inflation rates in both periods. Relative to this option, dropping non-rental housing services from the CPI considerably reduces the weight this sector receives in the price index and leads to a sizable bias in the measurement of inflation, whose sign depends on the evolution of the price of housing services relative to the CPI as a whole. During the 1985-2000 period, in Spain this bias has been approximately 0.35% per year in absolute terms.

VI. CONCLUSIONS

The treatment of non-rental housing in the CPI is an old topic in index number theory and practice. The issue requires a solution to two problems: (i) how to determine the weight that non-rental housing should have in the CPI, and (ii) how non-rental housing prices should be monitored over time. This paper is an empirical contribution to the first question using the 1980-81 and 1990-91 household budget surveys for Spain that have served to determine the official CPI weights for the 1983 and 1992 base, respectively.

The starting point is the economic approach to price index numbers for consumption goods and services, according to which a zero inflation rate means that the cost of a given standard of living has remained constant. As far as housing is concerned, the standard of living depends on the consumption of the housing services a dwelling provides regardless of the tenure mode in which the dwelling is held. The empirical problem addressed in this paper is how to impute a rental value to the flow of housing services provided by the non-rental housing stock in the base period. The CPI weight for such services will then be equal to the ratio of the imputed value to the total household expenditures in all commodities (including the non-rental housing services in question).

From a methodological perspective, the hedonic imputation method advocated in this paper has introduced two improvements upon the traditional procedure. First, Multiple Correspondence Analysis is used to select the linear combination of physical attributes that accounts for the maximum variance in this multiple-dimensional space for the entire housing stock. It has been shown that this scalar index can be interpreted as combining those aspects of the physical attributes that contribute to housing quality. It is claimed that using this housing quality index in the explanation of market rents presents several advantages over the usual alternative where all physical attributes are included as separate regressors. Second, this paper has recognized the possible interdependence between the choices of tenure mode and housing characteristics. To correct for the possible selection bias caused by systematic differences in housing characteristics between the market rental sector and the non-rental housing stock, a Heckman procedure has been used for the first time in this context. In this way, the estimated coefficients of the variables explaining market rents can be safely used to impute a rental value to non-rental housing units.

The more remarkable result of the paper is that the hedonic values thus obtained through an “objective” statistical procedure are not that different from the self-imputed values “subjectively” selected by the occupants. Hedonic values are less dispersed and, on average, 10% higher and 15% lower than the self-imputed ones in 1980-81 and 1990-91, respectively. Such differences do not lead to important departures in average annual inflation rates for the 1985-2000 period.

Fully liberalized and large rental housing markets are complex institutions where a highly differentiated commodity is rented for a price summarizing the value of the services provided by a large set of housing characteristics. In the Spanish case, the market rental sector only represents 8.9% and 5.3% of permanently occupied residential housing in 1980-81 and 1990-91, respectively. This small rental sector, which is still partly regulated, works under the influence of a rent controlled and a publicly subsidized rental sector. On the other side of the spectrum, a dominant non-rental sector represents 71.1% and 85.0% of the housing stock in 1980-81 and 1990-91, respectively. Under these conditions it would appear that, whatever its conceptual merits, the rental equivalence approach to the determination of a CPI weight to non-rental housing services is bound to fail in practice. However, the experiment performed in this paper has shown that two radically different imputation procedures lead to very close inflation rates during a 15 year period. Thus, in spite of a thin market rental sector and against all odds, occupants of non-rental housing in Spain and hedonic procedures arrive, on average, to comparable imputed values.

What is the explanation? On the statistical side, it should be emphasized that hedonic methods to explain rents and prices in rental and stock housing markets have been working reasonably well for 30 years in a variety of institutional scenarios. As this paper has shown, what is needed for these methods to also work in countries like Spain is simply large samples and good information on housing characteristics, legal status and year of occupancy. On the other hand, it seems that people are keenly aware of how housing markets work, so that self-imputed rental values – although more dispersed than hedonic ones- do contain useful information.

Finally, it should be emphasized that the aim of the exercise is simply to assess the weight that a given commodity should be assigned in a large commodity space. Tolerable differences in the numerator of the expression for determining that weight need not lead to significant differences in the magnitude of the weight itself. From this point of view, it is not so surprising that the two methods studied in this paper – or a combination thereof that has not been attempted here- lead to similar solutions in terms of inflation rates for the economy as a whole. This means that even in *a priori* unfavorable circumstances, like those of Spain, as long as good data is available it should be quite possible to find an acceptable solution to the problem at hand.

There is no room in this paper for a proper discussion of how the prices of non-rental housing services should be monitored over time. However, the treatment of quality change, as well as depreciation and aging bias in both rental and non-rental housing sectors is bound to rely on hedonic methods (see, *inter alia*, Randolph 1988, Hoffman and Kurz 2001, and Silber 1999).[35] The results of this paper, that add to a successful literature, show that these methods are worth pursuing.

This paper contains a warning for the specialists in the housing sector, the officials in charge of these matters, and public opinion in general: in a country like Spain with a large non-rental housing stock, eliminating its services from the CPI has been estimated to give rise to a 0.35% per year bias in the measurement of inflation. Given the alternatives, this is an unnecessarily crude form of dealing with a difficult problem. It appears that there is room for national and international statistical offices to keep experimenting within the sound economic approach assumed in this paper. It might not be otiose to end insisting that, together with good data, the successful development of this approach requires a staff capable of applying multivariate statistical techniques.

APPENDIX

This Appendix has two parts. Part A is devoted to a summary of the essentials of Multiple Correspondence Analysis, or MCA for short. MCA is a descriptive technique for representing contingency tables, that is, tables consisting of the frequencies with which a set of values of two or more qualitative variables appear in a data set (see Tenenhause and Young 1985, and Greenacre 1984 for a detailed treatment of MCA). Part B contains an application of MCA to the data on housing attributes coming from the 1980-81 and 1990-91 EPFs.

A. A Description of MCA

[35] This is also the case in the more promising version of the user cost approach as exemplified in Linneman and Voight (1991) and Crone *et al.* (2000).

Assume there are N observations on housing units, each one characterized by Q physical attributes, indexed by $q = 1,\ldots, Q$, and possibly correlated. Let J_q be the number of categories or modalities of variable q, and let $J = \Sigma_q J_q$ be the total number of categories with $N >> J$. The data can be represented by a (N x J) matrix Z, whose element z_{ij} takes the value 1 when housing unit i has modality j and zero otherwise, $i = 1,\ldots, N$ and $j = 1,\ldots, J$. Note that since each housing unit can only have one of the J_q categories of variable q, the row sums in Z must all be equal to Q, i. e. $\Sigma_j z_{ij} = Q$ for each $i = 1,\ldots, N$. Denote by N_j the absolute frequency of category j, or the sum of elements in column j, $\Sigma_i z_{ij} = N_j$, so that for each variable q, $\Sigma_{j\in I_q} \Sigma_i z_{ij} = N$. Finally, $\Sigma_i \Sigma_j z_{ij} = N Q$.

The objective of MCA is to obtain a set of uncorrelated variables w_k, indexed by $k =1,\ldots, K$, with $K < Q$, where each w_k is a linear combination of the J categories. In other words, the objective of MCA is to find a (J x K) matrix M where m_{jk} is the contribution of the j-th category to the new variable w_k, so that the information in the original data can be expressed through the lower-dimensionality (N x K) matrix W defined by

$$W = Z M,$$

whose columns are the w_k variables. As will be seen in Part B of this Appendix, the first variable w_1, which will be interpreted as a housing quality index, explains a large part of the variance in the data.

The matrix M is constructed as follows. Denote by F the relative frequency matrix, i. e. $F = (1/N Q) Z$. The average column profile is the (N x 1) vector $r = F 1_J$ with $r_i = 1/N$ for all $i = 1,\ldots, N$, while the average row profile is the (J x 1) vector $c = F^T 1_N$ with $c_j = (N_j / N Q)$ for each $j = 1,\ldots, J$. The corresponding (N x N) and (J x J) diagonal matrices are denoted D_r and D_c, respectively. Define the (N x J) matrix E by

$$E = D_r^{-1/2} (F - r\, c^T)\, D_c^{-1/2}.$$

The element $e_j^T e_j$ in the diagonal of the (J x J) matrix $E^T E$ is the χ^2 distance between the j-th column profile in matrix F and the average column profile r, weighted by its relative frequency c_j.[36] The sum is called the total inertia, TI, and it can be shown to be equal to $(J/Q) - 1$, i. e. $TI = \Sigma_j e_j^T e_j = (J/Q) - 1$.

MCA computes the singular value decomposition of E, say $U D_\alpha V^T$, where U and V are orthogonal matrices, and D_α has $J - Q$ non-zero eigenvalues. In practice, since $N >> J$, it is more convenient to compute the singular value decomposition of the diagonal matrix $E^T E$, say $E^T E = \Gamma D_\lambda \Gamma^T$, where $D_\lambda = D_\alpha^2$. The eigenvalues of D_λ quantify the inertia projected through each of the associated eigenvectors (which form the columns of Γ). The eigenvectors represent orthogonal directions of projection of centered column profiles. The direction of the first eigenvector associated to the largest eigenvalue is the optimal projection in the sense that it is the linear orientation that collects the maximum disparity between individuals according to the Q variables. The second eigenvector is orthogonal to the first one and represents the linear orientation that captures the maximum residual disparity not taken into account by the first projection axis. The remaining eigenvectors can be similarly interpreted until the total inertia is accounted for by K orthogonal axis with $K \leq J - Q$. In geometric terms, a change of axis is being performed where the original space of profiles that has dimension $J - Q$ is projected in a reduced space with dimension K. The coordinates used in the projection are contained in the (J x K) matrix M defined by

$$M = D_c^{-1/2}\, \Gamma\, D_\lambda^{1/2}.$$

For each k, the eigenvalue λ_k represents a percentage of the total inertia, but these percentages tend to be small and show a pessimistic idea about the proportion of the projected inertia by each axis. Thus, Benzécri (1979) proposes considering

[36] There is an analogous interpretation for the elements $e_j e_j^T$ in the diagonal of the (J x J) matrix $E E^T$ that refers to the rows.

solely the relevant P axis, that is, the axis associated with those eigenvalues with $\lambda_p > 1/Q$, $p = 1,\ldots, P$, and $P \leq K$. Analogously, he proposes to correct the eigenvectors with the transformation

$$\lambda_p^c = [Q/(Q-1)]^2\,[\lambda_p - (1/Q)]^2$$

and show the proportion of inertia explained in relation to $\Sigma_p \lambda_p^c$. In this way, the dimensionality of the original matrix is reduced from J – Q categories to P indicator variables losing a small quantity of information. Once the importance of each indicator is evaluated in this manner, its interpretation can proceed in terms of its correlation with all initial variables and the weight that each category receives in the corresponding linear combination.

B. An Index of Housing Quality

In 1980-81, there are Q = 8 qualitative variables and J = 23 categories, while in 1990-91 Q = 18 and J = 52. The number of observations with complete information along these dimensions is 23,898 in 1980-81 and 20,799 in 1990-91[37], representing a population of 9,992,051 and 11,105,215 housing units, respectively. (See Table A for the frequency distribution of all physical attributes in both years).

The percentage of inertia accounted for by the first two factors is 86.2% and 12.1% in 1980-81, and 73.2% and 10.5% in 1990-91.Moreover, variables with two or more categories show a parabolic structure when depicted in the plane defined by those two factors. This is a very frequent phenomenon, known as the *Guttman effect* (see Grenacre 1991), which simplifies the interpretation. Indeed, this effect reveals that whereas the first factor summarizes the order structure of all modalities, the second factor shows an opposition between extreme categories (low frequency) and average ones (large frequency) of a variable. Thus, in the sequel the analysis will continue in terms of the first factor that contains the relevant information about the variability among housing units as far as their physical attributes is concerned.

[37] In 1980-81 there are 73 observations without information on building age, while in 1990-91 this number is 289, plus 67 dwellings without information on housing size in squared meters.

Relative to the initial number of variables and categories, this constitutes a very drastic simplification indeed.

The correlation between the first factor and each of the Q physical attributes for both years is shown in Table B. Among the 8 common variables, hygienic services and water facilities are the most influential. Telephone, building age and heating facilities occupy an intermediate position above garage and housing size, while having electricity is the most prevalent characteristic in both years and the one with the smallest correlation with the index whose meaning is being discussed. Except hygienic services and water facilities, the remaining 6 common variables gain some importance in the determination of the index in 1990-91; for instance, the correlation coefficient of garage increases from 0.41 in 1980-81 to 0.58 in 1990-91. Nevertheless, having an elevator or air conditioning inside the house, or having sport facilities and other community services around it, are the characteristics with the largest correlation coefficients with the index in 1990-91.

Table B around here

In each year, each modality j receives a certain weight m_j, which can be positive or negative, but the more frequent a modality j is, the closer to zero m_j will be. Table C presents the rankings of modalities in both years in terms of normalized weights m_j' that preserve the sign and the ratios of the original weights. The normalization consists of assigning the values – 10 and 10 to the modalities with extreme m_js according to the formula

$$m_j' = m_j\,(10 / m^*),$$

where $m^* = \max \{| m_1|,\ldots, | m_J|\}$. In both years, the maximum normalized weight is below 10, indicating that there is no modality that plays the exact opposite role to the one that reaches the maximum negative value – 10.

Table C around here

As far as the interpretation of the index is concerned, it is very revealing that for all variables with two categories, the worst one receives a negative weight while the best one receives a positive one. For the remaining variables with 3 or more categories, weights are always naturally ordered from worst to best modality. In particular, the categories with the maximum negative influence in both years are having no hygienic services, no water (or only cold water), or no light. Moreover, the first two are highly correlated with the index (see Table B), which indicates that both of them are very influential in the interpretation. Among the remaining common variables in both years, having a telephone, being in a building less than 11 years old, and having a garage or more than a full bathroom are categories associated with high and increasing positive weights. In 1990-91, where there are 10 more variables, having a swimming pool, elevator, electric energy or natural gas for cooking, and a sports area are the modalities exercising the maximum positive weight in the index. Other aspects worth noticing when comparing the two years are as follows:

1) In 1980-81, to have any kind of heating system has a weight of 2.79 and not to have it detracts only – 1.73. In 1990-91, the range of variation goes from only 0.22 to – 3.62. Moreover, the correlation between the variable and the index goes from 0.59 in 1980-81 to 0.65 in 1990-91. The interpretation is that “heating” is a less frequent commodity in 1980-81 than in 1990-91, so that its possession leads to a high relative weight, while its absence is the feature that makes its presence felt more forcefully in 1990-91.

2) Having or not having a telephone plays the same pattern in both years: in 1980-81 the corresponding weights are 2.29 and – 2.17, while in 1990-91 the weights are 1.14 and – 3.27.

3) In both years, having a garage plays a stronger positive role than not having it, but more so in 1980-81 (3.47 *versus* – 0.68) than in 1990-91 (2.41 *versus* – 1.02). However, the correlation coefficient between this variable and the index increases from 0.41 in 1980-81 to 0.58 in 1990-91.

In our opinion, both the correlation between the index and the different variables and the way the weights received by worst and best modalities are naturally ordered for all variables without exception, indicate that the first factor obtained from MCA can be safely interpreted as an index of housing quality. The index values assigned to each housing unit as a function of its specific categories and the weights just analyzed, indicates the positive or negative deviation relative to the quality attributed to the "average housing unit" with the more frequent categories, whose index value is zero by construction. The common categories associated to the average housing unit in both years are the following: building age between 20 and 30 years, presence of electricity facilities, one bathroom (although perhaps not a full one), hot water, between 61 and 90 m^2, and without a garage. The average dwelling in 1980-81 does not possess telephone nor heating, while in 1990-91 it has both and it is situated in buildings with more than 3 floors, without an elevator, air conditioning, garden, swimming pool or sports area; gas is the fuel used for cooking and heating water, while for the remaining heating uses –served by mobile tools- this fuel is combined with electric energy.

The housing quality index has, of course, several shortcomings. First, the index does not take into account the heterogeneity in the way that certain categories influence quality for different housing types. For instance, heating in the Canary Islands or Andalucía should receive a smaller weight on quality than in the North of the country. Although such differences could be recognized by computing a different index for each housing type, the resulting quality indexes will not be comparable across types. Thinking of the main aim of this paper, a single housing quality index has been computed.[38] Second, the housing quality concept is restricted to the Q variables for which there is information in the EPFs. Judging from the existing hedonic literature on housing, there are several quality dimensions that are potentially important, such as how well preserved dwellings are, as well as basic neighborhood characteristics like pollution, safety, public transport facilities, and distance to the main centers of economic and recreational activity.

In any case, the real issue is whether, with all its shortcomings, the housing quality index here constructed behaves well in explaining market rents –an issue covered in Section III of the paper.

[38] To limit the influence of this problem on the results, only the two categories "no heating" and "heating of any kind" have been considered.

REFERENCES

Arévalo, R. (2001), *El mercado de la vivienda en España*, unpublished Ph.D. dissertation, Universidad Complutense de Madrid.

Arnott, R. (1995), "Time for Revisionism in Rent Control?", *Journal of Economic Perspectives*, **9**: 99-120.

Balk, B. (1994), "The New Consumer Price Indices of Statistics Netherlands", *Statistical Journal of the United Nations*, **11**: 119-123.

Basu, K and P. Emerson (2000), "The Economics of Rent Control", *Economic Journal*, **110**: 939-962.

Benzécri J. (1979), "Sur le calcul des taux d'inertie dans l'analyse d'un questionnaire. Addendum et erratum", *Cahiers de l'Analyse des Données*, **4**: 377-378.

Börsch-Supan, A. (1986), "On the Wrest German Tenants Protection Legislation", *Journal of Institutional and Theoretical Economics*, **142**: 380-404.

Boskin, M., E. Dulberger, R. Gordon, Z. Griliches, and D. Jorgenson (1996), *Toward a More Accurate Measure of the Cost of Living*, Final Report, Senate Finance Committee.

Bover, O. (1993), "Un modelo empírico de la evolución de los precios de la vivienda en España (1976-1991)", *Investigaciones Económicas*, **XVII**: 65-86.

Castro, M.A. (1992), "El nuevo sistema de Indices de Precios de Consumo", *Situación*, BBV, **3-4**: 167-179.

Crone, T, L. Nakamura and R. Voith (2000), "Measuring Housing Services Inflation", *Journal of Economic and Social Measurement*, **26**: 153-171.

Dalen, J. (1999), "A Proposal for a New System of Aggregation in the Swedish Consumer Price Index", proceedings of the Otawa Group, Fifth Meeting, Reykjavík, Iceland, August 25-27.

Diewert, E. (2000), "The Consumer Price Index and Index Number Purpose", Department of Economics, University of British Columbia, Discussion Paper No. 00-02.

Eurostat, World Bank, International Monetary Fund, O.E.C.D., and United Nations (1993), *System of National Accounts*, United Nations, New York.

Francois, J. (1989), “Estimating Homeownership Costs: Owners’ Estimates of Implicit Rents and the Relative Importance of Rental Equivalence in the Consumer Price Index”, *American Real Estate Urban Economics Association Journal*, **17**: 87-99.

Gillingham, R. and W. Lane (1982), “Changing the Treatment of Shelter Costs for Homeowners in the CPI”, *Monthly Labor Review*, **105**: 9-14.

Goodhart, C. (2001), “What Weight Should Be Given to Asset Prices in the Measurement of Inflation?”, *Economic Journal*, **111**: F335-F356.

Greenacre, M.J. (1984), "*Theory and applications of correspondence analysis*", London: Academic Press.

Greenacre, M.J. (1991), “Interpreting Multiple Correspondence Analysis”, *Applied Stochastic Models and Data Analysis*, **7**: 195-210.

Guasch, J. L. and R. Marshall (1987), "A Theoretical and Empirical Analysis of the Length of Residency Discount in the Rental Housing Market", *Journal of Urban Economics*, **22**: 291-311.

Heckman, J. (1979), "Sample Selection Bias As A Specification Error", *Econometrica*, **47**: 153-161.

Hoffman, J. and C. Kurz (2002), “Rent Indices for Housing in West Germany, 1985 to 1989”, Discussion paper 01/02, Economic Research Center of the Deutsche Bundesbank.

Hill, P. (1999), “Inflation, the Cost of Living and the Domain of a Consumer Price Index”, prepared for the Conference of European Statisticians, Joint ECE/ILO Meeting on Consumer Price Indices, Geneva, November 24-27.

Hubert, F. (1995), “Contracting with Costly Tenants”, *Regional Science and Urban Economics*, **25**: 631-654.

INE (1982), *Censos de Población y Viviendas. Año 1981*. Instituto Nacional de Estadística, Madrid.

INE (1983), *Encuesta de Presupuestos Familiares. Metodología y resultados*. Instituto Nacional de Estadística, Madrid.

INE (1992a), *Encuesta de Presupuestos Familiares. Metodología*. Instituto Nacional de Estadística, Madrid.

INE (1992b), *Censos de Población y Viviendas. Año 1991*. Instituto Nacional de Estadística, Madrid.

Linneman, P. and R. Voith (1991), “Housing Price Functions and Ownership Capitalization Rates”, *Journal of Urban Economics*, **30**: 100-111.

Marshall, A. (1887), *Principles of Economics*, Fourth Edition, London: The Macmillan Co.

Miron, J.R. (1990), “Security of Tenure, Costly Tenants and Rent Regulation”, *Urban Studies*, **27**: 167-184.

Nagy, J. (1997), “Do Vacancy Decontrol Provisions Undo Rent Control?”, *Journal of Urban Economics*, **42**: 64-78.

Peña, D. and J. Ruiz-Castillo (1984), "Robust Methods of Building Regression Models. An Application to the Housing Sector", *Journal of Business and Economic Statistics*, **2**: 10-20.

Peña, D. and V. Yohai (1995), "The Detection of Influential Subsets in Linear Regression by Using an Influence Matrix", *Journal of the Royal Statistical Society B*, **57**: 145-156.

Quigley, J. (1979), “What Have We Learned About Urban Housing Markets?”, in P. Mieszkowski and M. Straszheim (eds.), *Current Issues in Urban Economics*, The Johns Hopkins University Press.

Randolph, W. (1988), “Housing Depreciation and Aging Bias in the Consumer Price Index”, *Journal of Business and Economic Statistics*, **6**: 359-371.

Rosen, S. (1974), “Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition”, *Journal of Political Economy*, **82**: 34-55.

Ruiz-Castillo, J., M. Izquierdo, and E. Ley (1999), *La medición de la inflación en España*, “la Caixa”, Barcelona.

Schlicht (1983), “The Tenant’s Decreasing Willingness To Pay and the Rent Abatement Phenomenon”, *Journal of Institutional and Theoretical Economics*, **139**: 155-159.

Sheppard, S. (1999), “Hedonic Analysis of Housing Markets”, in E. S. Mills and P. Chesire (eds.), *Handbook of Regional and Urban Economics*, Elsvier Science.

Silver, M. (1999), “An Evaluation of the Use of Hedonic Regressions for Basic Components of Consumer Price Indices”, *Review of Income and Wealth*, **45**: 41-56.

Smith, L, K. Rosen, and G. Fallis (1988), "Recent Developments in Economic Models of Housing Markets", *Journal of Economic Literature*, **XXVI**: 29-64.

Tenenhaus, M. y Young, F.W. (1985), "An analysis and synthesis of multiple correspondence analysis, optimal scaling, dual scaling, homogeneity analysis and other methods of quantifying categorical multivariate data", *Psychometrika*, **50**: 91-119.

Triplett, J. (1990), "Hedonic Methods in Statistical Agency Environments: An Intellectual Biopsy", in E. Berndt and J. Triplett (eds.), *Fifty Years of Economic Measurement*, Chicago: University of Chicago Press.

Triplett, J. (2001), "Should the Cost-of-living Index Provide the Conceptual Framework for a Consumer Price Index?", *Economic Journal*, **111**: F311-F334.

Turvey, R., Sellwood, D., Szuke, B., Donkers, H. *et al.*, Marret, M., Clements, L. *et al.*, Woodhouse, T., and Hanson, K. (1989), *Consumer Price Indexes*, Geneva: International Labour Organization.

Turvey, R. (1999), "True Cost of Living Indices", proceedings of the Otawa Group, Fifth Meeting, Reykjavík, Iceland, August 25-27.

Turvey, R. (2000), "Owner-occupiers and the Price Index", *World Economics*, **1**: 153-159.

United States Department of Labor, Bureau of labor Statistics (1997), Chapter 5, "Published Estimates of Consumer Price Elasticities of Demand", in *The Use of the Geometric Mean in the Elementary Indexes of the Consumer Price Index*, unpublished study quoted in Triplett (2001).

Table 1. Permanent Residential Housing by Tenure Mode, and Secondary Housing. A Comparison of the EPFs with the Censuses (In Thousand Units)

RESIDENTIAL HOUSING	1980-81 EPF	%	1981 Census	%
A. Permanent Housing	**10,022**	**100.0**	**10,431**	**100.0**
I. Rental Housing	2,297	22.9	2,169	20.8
II. Owner-occupied	6,928	69.1	7,629	73.1
Owned Outright	5,380	53.7	5,764	55.2
Being Paid	1,548	15.4	1,865	17.9
III. Other	797	8.0	633	6.1
Wages in Kind	219	2.2	322	3.1
Transferred	578	5.8	311	3.0
B. Secondary Housing	**666**	**-**	**1,900**	**-**

RESIDENTIAL HOUSING	1990-91 EPF	%	1991 Census	%
A. Permanent Housing	**11,298**	**100.0**	**11,736**	**100.0**
I. Rental Housing	1,694	15.0	1,781	15.2
II. Owner-occupied	8,789	77.8	9,194	78.3
Owned Outright	7,504	66.4	7,361	62.7
Being Paid	1,285	11.4	1,833	15.6
III. Other	815	7.2	761	6.5
Wages in Kind	134	1.2	189	1.6
Transferred	681	6.0	572	4.9
B. Secondary Housing	**1,284**	**-**	**2,923**	**-**

Table 2. Permanent Residential Housing by Tenure Mode and Legal Status in the EPFs.
Sample and Population Statistics

	1980-81 EPF			1990-91 EPF		
SECTOR	**Sample**	**Population**	**%**	**Sample**	**Population**	**%**
I. RENTALHOUSING*	**5,484**	**2,297,105**	**22.9**	**2,975**	**1,694,184**	**15.0**
1. Market Sector	2,181	888,945	8.9	1,061	601,970	5.3
2. Rent Controlled	968	405,290	4.0	229	129,968	1.1
3. Public Housing	1,787	773,164	7.7	771	426,875	3.9
4. Unknown Legal Status	535	223,795	2.2	914	535,371	4.7
II. OWNER-OCCUPIED	**16,427**	**6,928,150**	**69.1**	**16,623**	**8,789,287**	**77.8**
5. Market Sector	9,307	4,104,814	41.0	9,132	4,883,659	43.2
6. Public Housing	5,316	2,048,206	20.4	5,222	2,676,098	23.7
7. Unknown Legal Status	1,804	775,130	7.7	2,269	1,229,530	10.9
III. OTHER	**2,060**	**798,911**	**8.0**	**1,557**	**815,038**	**7.2**
TOTAL	**23,971**	**10,024,166**	**100.0**	**21,155**	**11,298,509**	**100.0**

* This sector includes 13 sample observations, representative of 5,911 housing units at the population level, that cannot be classified in the market or the rent controlled sector because they lack information on the year of first occupancy.

Table 3. Frequency Distribution and Mean Monthly Rent (in euros) for Physical Attributes, Geographic Characteristics, Years of Occupancy and Quality Index In The Market Rental Sector According to the 1980-81 EPF

Physical Attributes	Population Distribution	Mean Rent	Years of Occupancy	Population Distribution	Mean Rent	
Sample Size	**867,627 = 100.0%**	**33.8**				
1. Hygienic Services						
Hyg1 = Less than a full bathroom	18.7	13.6	1. Ocup65	4.0	17.6	
Hyg2 = One full bathroom	69.1	33.3	2. Ocup66	2.4	15.4	
Hyg3 = More than a full bathroom	12.2	67.7	3. Ocup67	2.3	15.7	
2. Water Facilities			4. Ocup68	4.6	17.4	
Water1 = No water	3.9	11.7	5. Ocup69	3.9	20.8	
Water2 = Only cold water	23.8	16.9	6. Ocup70	7.5	24.0	
Water3 = Individual hot water		68.1	37.8	7. Ocup71	3.9	17.9
Water4 = Centrally heated hot water	**4.2**	**85.0**	**8. Ocup72**	**7.5**	**23.3**	
3. Heating Facilities			9. Ocup73	7.2	22.1	
Heat1 = No heating	66.3	26.0	10. Ocup74	10.0	26.2	
Heat2 = Only mobile tools	19.9	34.3	11. Ocup75	8.7	29.9	
Heat3 = Fixed heating installation	13.8	70.6	12. Ocup76	6.6	35.7	
Heat12 = Heat1 and Heat2	86.2	28.0	13. Ocup77	6.4	35.7	
4. Garage Facilities			14. Ocup78	7.6	57.1	
No garage facilities	88.3	31.1	15. Ocup79	9.8	65.7	
Garage	11.7	54.4	16. Ocup80-81	7.7	56.5	
5. Telephone Facilities						
No telephone	63.1	27.1	Ocup6573	43.3	20.5	
Telephone	36.9	45.3	Ocup7477	31.6	31.1	
6. Housing Size			Ocup7880	25.1	60.2	
Size1 = Less than 60 m²	18.7	22.8				
Size2 = 61 – 90	46.7	32.6	Ocup7480	56.7	44.0	
Size3 = 91 - 130	27.9	38.2				
Size4 = More than 130 m²	6.7	55.0				
7. Building Age						
Age1 = More than 50 years	26.6	19.2				
Age2 = 31 – 50	16.0	25.8				
Age3 = 21 - 30	10.6	31.3				
Age4 = 11 – 20	28.6	36.2				
Age5 = Less than 10 years	18.2	59.9				

Geographical Char.	Population Distribution	Mean Rent	Qindex		Mean Rent
1. Municipal Size			**Quartiles**		
Mun1 = Less than 2,000	7.6	19.9	I	-0.72	17.3
Mun2 = 2,001 - 10,000	14.7	20.6	II	-0.12	27.1
Mun3 = 10,001 – 50,000	20.7	25.0	III	0.22	40.7
Mun4 = 50,001 – 500,000	34.8	36.6	IV	0.65	71.3
Mun5 = More than 500,000	22.2	51.3			
Mun45 = Mun4 + Mun5	57.0	42.3	Qindex6573	-0.15	20.5
2. Mean Prov. Housing Price			Qindex7480	-0.09	44.0
Prov1[1] = Below 860 euros / m^2		5.5	26.8		
Prov2[2] = 861 - 1,230 euros / m^2		40.8	26.5		
Prov3[3] = 1,231 – 1,700 euros / m^2	19.0	33.1			
Prov4[4] = Above 1,700 euros / m^2	34.7	44.0			

Prov123 = Prov1 + Prov2 + Prov3	65.3	28.4

Notes:

[1] Cáceres, Murcia, Badajoz, Lugo, Ceuta, Melilla

[2] Cádiz, Girona, Málaga, Segovia, Granada, Palencia, León, A Coruña, Almería, Soria, Toledo, Tarragona, Ávila, Guadalajara, Huelva, Valencia, Castellón, Huesca, Córdoba, Ourense, Alicante, Ciudad Real, Lleida, Cuenca, Albacete, Zamora, Teruel, Jaén, Pontevedra, Santa Cruz de Tenerife

[3] Burgos, Palma de Mallorca, Santander, Salamanca, Logroño, Las Palmas de Gran Canaria, Valladolid, Zaragoza, Oviedo, Sevilla

[4] Madrid, Barcelona, San Sebastián, Bilbao, Vitoria, Pamplona

Table 4 Regression Results. Classical Hedonic Models, 1980-81

	MODEL I		**MODEL II**		**MODEL III**		**MODEL IV**	
Variables	**Coeff.**	**t**	**Coeff.**	**t**	**Coeff.**	**t**	**Coeff.**	**t**
Constant	2.1472	7.8	2.5537	9.8	2.7163	10.3	2.7772	11.1
Hyg1	- 0.2912	- 4.8	-0.2410	- 4.5	- 0.2427	- 4.4	- 0.2457	- 5.0
Hyg3	0.4259	6.3	0.3650	6.5	0.3751	6.7	0.3631	7.2
Water1	- 0.5689	- 4.6	- 0.5127	- 4.1	-0.4975	- 3.9	- 0.5101	- 4.3
Water2	- 0.2506	- 4.7	- 0.1847	- 3.8	- 0.1824	- 3.7	- 0.2340	- 5.2
Water4	**0.4851**	**3.9**	**0.3781**	**3.5**	**0.3503**	**3.2**	**0.2883**	**2.9**
Heat3	0.2871	3.4	0.2664	4.1	0.2578	3.9	0.2280	3.9
Garage	0.1303	1.7	0.0934	1.4	0.0874	1.3	0.1422	2.6
Phone	- 0.0243	- 0.5	0.1221	2.9	0.1209	2.9	0.1128	3.0
Lnsize	0.2494	4.2	0.2287	4.1	0.2388	4.2	0.2354	4.3
Age	-0.0039	- 6.2	- 0.0032	- 5.5	- 0.0032	- 5.4	- 0.0030	- 5.7
Mun1	-0.4322	- 5.4	- 0.3835	- 4.5	- 0.3868	- 4.5	- 0.4073	- 5.0
Mun2	-0.3492	- 5.8	- 0.2825	- 5.4	- 0.2856	- 5.4	- 0.2624	- 5.8
Mun3	-0.3390	- 6.2	-0.3139	- 6.7	- 0.3131	- 6.8	- 0.2389	- 6.5
Prov4	0.2598	5.4	0.2633	6.3	0.2708	6.4	0.2723	7.2
Ocup65			- 0.9766	- 7.4				
Ocup66			-1.2100	- 9.9				
Ocup67			-1.1187	- 7.9				
Ocup68			- 1.0605	- 9.8				
Ocup69			- 0.8724	- 7.9				
Ocup70			- 0.7895	- 7.6				
Ocup71			- 0.8587	- 9.1				
Ocup72			- 0.8067	- 8.9				
Ocup73			- 0.7992	- 8.6				
Ocup74			- 0.6336	- 7.7				
Ocup75			- 0.4798	- 5.6				
Ocup76			- 0.3024	- 3.5				
Ocup77			- 0.2969	- 3.3				
Ocup78			- 0.2100	- 2.1				
Ocup79			0.0812	1.0				
Years6573					- 0.0400	- 3.5	- 0.0392	- 3.9
Years7480					- 0.1149	-10.7	- 0.1275	-13.2
Ocup6573					- 0.6450	- 4.7	- 0.7081	- 5.9

n	2,142	2,142	2,142	2,052
Root MSE	0.7784	0.6916	0.6935	0.5799
R^2	0.361	0.499	0.493	0.586

Table 5. Regression Results. Hedonic Models with the Housing Quality Index, 1980-81

Variables	**MODEL A** Coeff.	t	**MODEL B** Coeff.	t	**MODEL C** Coeff.	t
Constant	2.0187	18.6	1.7957	15.0	1.8974	17.5
Mun1	-0.3702	- 4.6	-0.3779	- 4.6	-0.3783	- 4.9
Mun2	-0.2844	- 5.3	-0.2876	- 5.4	-0.2558	- 5.7
Mun3	-0.3078	- 6.5	-0.3042	- 6.5	-0.2245	- 6.1
Prov4	0.2358	5.7	0.2367	5.7	0.2375	6.4
Years6573	-0.0394	- 3.4	-0.0414	- 3.5	-0.0405	- 4.0
Years7480	-0.1180	-11.1	-0.1151	-10.9	-0.1281	-13.3
Ocup6573	-0.6937	- 5.0	-0.1815	- 0.8	-0.3786	- 1.9
Qindex	0.7925	21.2				
Qindex6573			0.6789	11.2	0.7253	14.7
Qindex7480			0.8800	20.2	0.8717	22.2
n	2,142		2,142		2,052	
Root MSE	0.6994		0.6977		0.5844	
R^2	0.482		0.485		0.578	

Table 6. Frequency Distribution and Mean Monthly Rent (in euros) for Physical Attributes, Characteristics, and Years of Occupancy and Quality Index In The Market Rental Sector According to the 1990-91 EPF

Physical Attributes	Population Distribution	Mean Rent		Population Distribution	Mean Rent
1. Hygienic Services			**10. Water Heating Fuel**		
None or Outside the Dwelling	3,0	53,0	None	8,6	37,6
One Full Bathroom the Dwelling	79,3	89,6	Solid (Wood, Coal, etc.)	1,9	71,9
More than a Full Bathroom	17,8	171,3	Gaseous (Butane, Propane)	58,4	97,8
2. Water Facilities			Liquid (fuel-oil)	4,1	212,0
No Water	0,45	13,7	Electric Energy	18,4	90,4
Only Cold Water	8,11	38,9	Natural Gas	8,6	185,0
Individual Hot Water	85,16	104,6	**11. Heating System Fuel**		
Centrally Heated Hot Water	6,28	171,7	None	13,7	109,2
3. Heating Facilities			Solid (Wood, Coal, etc.)	12,8	77,7
No Heating or only Mobile Tools	77,9	85,3	Gases (Butane, Propane)	22,8	79,1
Private Heating	15,6	156,7	Electric Energy	41,6	95,8
Central Heating	6,5	187,9	Liquid (gas oleo, Fuel-oil)	6,2	192,8
4. Garage Facilities			Natural Gas	3,0	283,4
No Garage Facilities	81,2	92,8	**12. Cooking Fuel**		
Garage	18,9	147,2	Solid (Wood, Coal, etc.)	1,4	43,8
5. Telephone Facilities			Gaseous (Butane, Propane)	81,8	91,8
No Telephone	35,6	84,4	Electric Energy	7,4	155,8
Telephone	64,4	113,4	Natural Gas	9,4	168,2
6. Housing Size			**13. Garden Facilities**		
Less than 60 m2	18,9	85,8	None	91,5	101,8
61-90	38,9	99,1	Garden	8,6	116,2
91-130	30,1	105,4	**14. Building Type**		
More than 130 m2	12,2	136,5	Inferior Tenement	0,4	71,2
7. Building Age			One Floor	18,7	63,0
More than 50 years	29,3	81,4	Two Floors	6,4	75,3
31-50	9,0	81,4	Three Floors or more	74,6	115,6
21-30	26,5	87,3	15 Swimming Pool		
11-20	27,8	133,9	None	98,1	102,4
Less than 10 years	7,4	155,6	Swimming Pool	1,9	135,5
8. Elevator Facilities			**17. Sports Area**		
None	73,4	77,1	None	98,9	102,9
Elevator	26,6	174,6	Sports Area	1,1	112,1
9. Air Conditioning			**18.Other Community Services**		
None	98,2	102,2	None	51,1	69,9
Air Conditioning	1,8	151,5	Some Community Services	48,9	137,7

Geographical Char.	Population Distribution	Mean Rent
1. Municipal Size		
Mun1=Less than 2,000	3,0	53,0
Mun2=2,001-10,000	13,3	74,4
Mun3=10,001-50,000	23,6	88,5
Mun4=50,000-500,000	35,8	107,4
Mun5=More than 500,000	24,2	132,6
2. Mean Prov. Housing Price		
Prov1=Below 860 euros / m2	5,2	62,8
Prov2=861-1,230 euros / m2	36,1	99,6
Prov3=1,231-1,700 euros / m2	20,2	101,0
Prov4=Above 1,700 euros / m2	38,5	112,7

Years of Occupancy	Population Distribution	Mean Rent
Ocup6575	29,20	6.796
Ocup7682	24,80	17.404
Ocup8390	46,00	23.575
Qindex Quartiles	**Mean Qindex**	**Mean Rent**
I	-5,14	56,6
II	-1,80	84,6
III	0,76	118,7
IV	6,24	176,8
Qindex6575	-1,70	40,8
Qindex7682	-0,27	104,6
Qind8390	0,05	141,7

Table 7. Regression Results. Hedonic Models with the Housing Quality Index, 1990-91

Variables	MODEL A Coeff.	t	MODEL B Coeff.	t
Constant	3.6176	21.4	3.9188	26.1
Mun1	-0.3993	-2.4	-0.3762	-3.2
Mun2	-0.1965	-1.9	-0.2491	-2.9
Mun3	-0.1483	-2.2	-0.1896	-2.9
Prov1	-0.3453	-3.1	-0.3429	-3.9
Years6575	-0.0747	-3.9	-0.0656	-3.5
Years7690	-0.0056	-0.4	-0.0270	-2.2
Ocup6575	-0.4605	-1.0	-0.8752	-2.2
Ocup7682	-0.7911	-2.4	-0.7929	-2.5
Qindex6575	0.1220	8.3	0.1210	9.2
Qindex7682	0.0987	4.9	0.0969	4.9
Qindex8390	0.0754	9.0	0.0681	8.7
n	1,035		982	
Root MSE	0.7545		0.6364	
R^2	0.5070		0.5921	

Table 8. Accumulated Discounts By Year of Occupancy as a Percentage of the Rent Paid In the Sample Year for a Dwelling of Average Quality, and Official Housing Inflation Rate (CPI Based in 1992)

Accumulated Discount, In %

	1980-81 EPF			1990-91 EPF			
Year of Occupancy	**Traditional Hedonic (1)**	**Qindex Regression Before Heckman Correction (2)**	**Qindex Regression After Heckman Correction (3)**	**Qindex Regression Before Heckman Correction (4)**	**Qindex Regression After Heckman Correction (5)**	**Housing Inflation Rate, In % (6)**	
1965	70.1	71.3	74.3	82.3	83.5	-	
1966	68.9	70.1	73.1	81.1	82.4	7.9	
1967	67.7	68.8	71.8		79.9	81.2	8.9
1968	66.4	67.5	70.4		78.5	79.9	6.8
1969	65.1	66.2	69.0		77.0	78.5	1.3
1970	63.7	64.8	67.5		75.5	77.0	4.87
1971	62.2	63.4	65.9		73.8	75.4	5.9
1972	60.7	61.8	64.3		72.0	73.7	5.1
1973	59.1	60.3	62.5		70.2	71.9	10.3
1974	53.5	53.6	56.3		68.1	70.0	12.9
1975	47.1	47.3	49.8		66.0	67.9	14.0
1976	40.0	40.1	42.4		51.6	54.6	11.0
1977	31.8	31.9	33.9		50.2	53.1	13.4
1978	22.5	22.6	24.1		48.9	51.6	12.6
1979	12.0	12.0	12.9		47.5	50.0	13.5
1980				46.0	48.3	12.1	
1981				44.6	46.6	13.3	
1982				43.0	44.8	12.1	
1983				17.2	20.3	8.6	
1984				15.0	17.7	5.6	
1985				12.6	15.0	6.7	
1986				10.2	12.2	7.5	
1987				7.8	9.3	5.2	
1988				5.3	6.3	6.5	

1989	2.7	3.2	9.1

Table 9. The Market Rental Sector *vs.* The Non-rental Housing Stock: Differences in Physical Attributes, Geographic Characteristics, and Other Characteristics

	1980-81		1990-91	
Physical Attributes	**Market Rental**	**Other Stock**	**Market Rental**	**Other Stock**
TOTAL	100.0	100.0	100.0	100.0
1. Hygienic Services*				
Hyg1 = No bathroom inside the building	4.9	7.3	7.4	4.1
Hyg2 = Less than one bathroom inside	13.9	8.0	74.1	68.9
Hyg3 = One o more than a bathroom	81.2	84.7	18.5	27.0
2. Water Facilities				
Water34 = Hot water	72.2	78.8	91.5	95.3
3. Garage Facilities				
Garage	11.3	18.6	19.7	30.4
4. Telephone Facilities				
Telephone	37.1	53.6	65.5	79.8
5. Housing Size				
Size1 = Less than 60 m^2	18.6	9.5	17.8	7.5
Size2 = 61-90	46.5	40.0	39.4	36.2
Size3 = 91-130	28.3	36.3	30.3	38.9
Size4 = Less than 60 m^2	6.6	14.2	12.5	17.4
6. Building Age				
Age1 = More than 30 years	46.6	34.8	42.3	26.8
Age2 = 11 – 30	41.0	41.7	53.1	58.6
Age3 = Less than 11 years	12.4	23.5	4..6	14.6
7. Building Type				
Detached = Detached House			18.1	38.9
8. Garden Facilities				
Garden			8.5	16.5
Geographical Characteristics				
1. Municipal Size				
Mun1 = Less than 2,000	7.6	12.9	2.9	8.3
Mun2 = 2,001 – 10,000	14.6	21.0	13.1	20.0
Mun3 = 10,001 – 50,000	20.5	21.0	23.9	22.6
Mun45 = More than 50,000	57.3	45.1	60.1	49.1
2. Mean Provincial Housing Price				
Prov1 = Below 860 euros / m^2	5.5	7.2	4.7	6.7
Prov2= 861- 1,230 euros / m^2	40.6	45.2	37.0	45.0
Prov3= 1,231- 1,700 euros / m^2	18.6	18.4	19.8	17.8
Prov4 = Above 1,700 euros / m^2	35.3	29.2	38.5	30.5
Other Characteristics				

1. Years of Occupation				
ocup1 = More than 30 years	0.0	14.5	0.0	13.8
ocup2 = 11-30	24.3	46.5	46.4	58.0
ocup3 = Less than 11 years	75.7	39.0	53.6	28.2
2. Migrant Household Head				
Migr75 = Arrived during 1975-1981	13.3	3.2		
Migr85 = Arrived during 1985-1991			12.9	3.4

***In 1990-91: Hyg1= Less than a full bathroom; Hyg2= One bathroom; Hyg3= More than one bathroom.**

Table 10. Hedonic Regression Models with a Selection Mechanism, 1980-81 and 1990-91

1980-81			1990-91		
Dependent variable: log rent			**Dependent variable: log rent**		
	Coeff.	t		Coeff.	t
Constant	2.0l35	16.6		3.8773	25.9
Mun1	-0.4250	- 5.2		-0.4106	- 3.4
Mun2	-0.2949	- 6.4		-0.2793	- 3.2
Mun3	-0.2480	- 6.6		-0.2033	- 3.2
Prov4	0.2520	6.8	Prov1	-0.3712	- 4.2
Years6573	-0.0471	- 4.4	Years6575	-0.0668	- 3.6
Years7480	-0.1380	-13.4	Years7690	-0.0325	- 2.5
Ocup6573	-0.4120	- 2.0	Ocup6575	-0.9067	- 2.2
Qindex6573	0.6496	11.2	Ocup7682	-0.7750	- 2.5
Qindex7480	0.7886	14.8	Qindex6575	0.1163	8.8
			Qindex7682	0.0926	4.6
			Qindex8390	0.0642	7.7

Selection Mechanism			**Selection Mechanism**		
	Coeff.	**t**		**Coeff.**	**t**
Constant	-153.63	-16.6	Constant	-38.1483	- 5.1
Hig1	-0.4688	- 4.9	Hig1	0.2149	1.7
Hig3	-0.1085	- 1.5	Hig3	- 0.1698	- 3.0
Water34	-0.2637	- 4.4	Garage	- 0.1100	- 1.9
Garage	-0.2271	- 3.1	Phone	- 0.4808	- 8.6
Phone	-0.4408	- 9.5	Size1	0.2787	3.6
Size*	-0.0038	- 6.1	Age***	- 0.0106	- 10.7
Age1	1.4250	24.0	Detached	- 0.4432	- 6.6
Age2	0.7870	14.4	Garden	- 0.0958	- 2.0
Ocup**	0.0778	16.6	Ocup	0.0294	7.8
Mun1	-0.5153	- 6.0	Mun1	- 0.3835	- 3.9
Mun2	-0.4473	- 6.9	Mun2	- 0.2255	- 3.5
Mun3	-0.2796	- 5.6	Prov1	- 0.0913	- 1.1
Prov1	-0.2570	- 2.8	Prov2	- 0.0370	- 0.8

Prov2	-0.1272	- 2.4	Migr85	0.4130	5.3
Prov3	-0.1154	- 2.1			
Migr75	0.5631	7.3			
r	0.214	2.8	r	0.149	1.4
Sigma	0.591	40.4	Sigma	0.637	25.7
n= 18,794			n= 17,804		

Wald test (ρ = 0): Chi2(1)=7.15; Prob > chi2 = 0.0075   Wald test (ρ = 0): Chi2(1)=1.85 Prob > chi2 = 0.1740

*** Size = m²**

**** Ocup = year of occupation**

*****Age = year of building**

Table 11. A Comparison of Hedonic and Self-imputed Rental Values for the Non-rental Housing Stock, 1980-81

TENURE MODE	ALL	Frequency Distribution of the Non-rental Housing Stock By Tenure Mode and Quartiles of the Distribution of Hedonic Values I	II	III	IV	
Owner-Occupied						
Market Sector	53.2	66.3	55.9	44.8	46.5	
Public Housing		26.5	3.6	19.7	37.9	42.7
Unknown Legal Status	10.0	15.7	10.9	8.1	6.0	
Other	10.3	14.4	13.6	9.2	4.8	
TOTAL	100.0	100.0	100.0	100.0	100.0	

Percentage of Units In Each Tenure Mode That Have Been Classified In The Same Quartile According To the Distributions Of Self-imputed And Hedonic Values, In %

TENURE MODE	I	II	III	IV	Spearman Coefficient	
Owner-Occupied						
Market Sector	62.9	38.5	37.3	61.6	0.72	
Public Housing		46.0	30.6	30.1	51.6	0.49
Unknown Legal Status	61.2	37.0	37.7	53.2	0.68	
Other	62.0	31.6	37.3	53.5	0.65	
ALL	63.1	37.0	36.2	57.3	0.66	

	Mean Monthly Rental Values (in euros)		Difference Between Hedonic and Self-imputed Mean Values, In %				
TENURE MODE	Hedonic	Self-imputed	Total	Quartiles Of The Hedonic Values			
				I	II	III	IV
Owner-Occupied							
Market Sector	54.2.	51.9	4.3	2.4	- 1.0	3.4	7.6
	(31.9)	(45.2)					
Public Housing	73.3	57.8	21.2	- 2.2	9.9	20.6	24.0
	(26.4)	(35.4)					
Unknown Legal Status	46.9	44.4	5.5	- 5.9	0.0	4.7	16.1
	(26.8)	(25.5)					
Other	46.1	41.5	10.0		6.0	4.5	14.8 13.4
	(25.1)	(34.3)					
ALL	57.7	51.6	10.6	6.1	2.7	10.1	14.1
	(31.0)	(41.1)					

Table 12. A Comparison of Hedonic and Self-imputed Rental Values for the Non-rental Housing Stock, 1990-91

TENURE MODE	ALL	Frequency Distribution Of The Non-rental Housing Stock By Tenure Mode and Quartiles of the Distribution of Hedonic Values				
		I	II	III	IV	
Owner-Occupied						
Market Sector	50.9	**67.4**	**59.0**	**41.7**	**38.6**	
Public Housing		27.9	**5.8**	**16.8**	**38.0**	**46.3**
Unknown Legal Status		12.8	**16.1**	**13.8**	**12.2**	**9.8**
Other		8.5	**10.7**	**10.4**	**8.1**	**5.3**
TOTAL	100.0	**100.0**	**100.0**	**100.0**	**100.0**	

TENURE MODE	Percentage of Units In Each Tenure Mode That Have Been Classified In The Same Quartile According To the Distributions Of Self-imputed And Hedonic Values, In %				Spearman Coefficient	
	I	II	III	IV		
Owner-Occupied						
Market Sector	**54.8**	**36.8**	**33.3**	**56.2**	**0.61**	
Public Housing		**41.9**	**27.4**	**30.2**	**47.4**	**0.46**
Unknown Legal Status		**55.3**	**34.2**	**33.9**	**53.1**	**0.60**
Other		**50.1**	**33.4**	**33.4**	**53.2**	**0.51**
ALL	54.9	32.1	30.0	53.5	0.57	

TENURE MODE	Mean Monthly Rental Values (in euros)			Difference Between Hedonic and Self-imputed Mean Values, In %					
	Hedonic			Self-imputed		Total		Quartiles Of The Hedonic Values	
					I	II	III	IV	
Owner-Occupied									
Market Sector	**147.5**	**184.8**		**- 25.2**	**- 11.1**	**- 24.8**	**- 29.9**	**- 29.4**	
	(70.1)	**(166.0)**							
Public Housing	**191.8**			**195.9**	**- 2.2**	**- 22.4**	**- 8.9**	**- 5.3**	**1.4**
	(63.5)	**(122.5)**							
Unknown Legal Status		**146.6**	**173.9**			**- 18.6**	**- 12.4**	**- 23.7**	
	(63.9)	**(127.2)**							
Other	**140.7**		**149.8**			**- 6.5**	**- 5.9**	**- 8.4 - 10.5**	

	(59.3)	**(128.8)**			
ALL	159.2	183.5	- 15.3	- 11.4	- 20.2
	(69.7)	(147.7)			

Table 13. Inter-annual Inflation Rates and Mean Annual Inflation Rate for 1985-1992 and 1993-2000 (Percentage Points Per Year)

Base = 1983	Treatment Of Non-rental Housing Services In The CPI:				
	Rental Equivalence Approach		Excluded From CPI		Non-rental Housing Prices
	Hedonic Imputations	Self-imputed Values		(4) =	
	(1)	(2)	(3)	(1) – (3)	(5)
SUBPERIODS					
August1985 - December1985	2.82	2.84	3.05	-0.24	0.62
1986	8.26	8.30	8.66	-0.40	4.47
1987	4.63	4.66	4.89	-0.26	2.09
1988	5.91	5.89	5.77	0.13	7.26
1989	6.92	6.91	6.84	0.08	7.76
1990	6.61	6.62	6.78	-0.17	4.95
1991	5.61	5.60	5.55	0.06	6.23
1992	5.41	5.44	5.69	-0.28	2.63
Mean Annual Rate,					
August 1985 – December 1992	9.26	9.29	9.58	-0.33	6.45

Base = 1992	Treatment Of Non-rental Housing Services In The CPI:

	Rental Equivalence Approach		Excluded From CPI		
	Hedonic Imputations	Self-imputed Values		(4) =	Non-rental Housing Prices
SUBPERIODS	(1)	(2)	(3)	(1) – (3)	(5)
January 1993 – December 1993	4.41	4.47	3.92	0.49	8.23
1994	4.40	4.41	4.33	0.07	4.92
1995	4.52	4.56	4.26	0.26	6.48
1996	3.78	3.85	3.23	0.55	7.85
1997	2.56	2.62	2.05	0.50	6.09
1998	1.76	1.80	1.39	0.36	4.20
1999	2.97	2.98	2.88	0.09	3.56
2000	3.99	3.99	4.00	-0.01	3.89
Mean Annual Rate,					
January 1993 – December 2000	4.19	4.24	3.81	0.38	7.19

Table A. Frequency Distribution for Physical Attributes, 1980-81 and 1990-91

Physical Attributes	1980-81	1990-91
Sample Size	**9,992,051 = 100.0%**	**11,105,215 = 100%**
1. Hygienic Services		
None or Outside the Dwelling	7.6	1.8
One Full bathroom Or Less	77.3	73.3
More Than A Full Bathroom	15.1	24.9
2. Water Facilities		
No Water	3.8	0.4
Only Cold Water	19.4	5.1
Cold and Hot Water	76.7	94.6
3. Heating Facilities		
No Heating	61.6	10.9
Mobile Tools, Ind. or Central Heating	38.4	89.1
4. Garage Facilities		
No garage	83.6	72.3
Garage	16.4	27.7
5. Telephone Facilities		
No telephone	49.3	22.7
Telephone	50.7	77.3
6. Electricity		
No Electricity	0.9	0.2
Electric Facilities	99.1	99.8
7. Housing Size		
Less than 60 m^2	12.7	9,0
61 – 90 m^2	41.6	37.8
91 – 130 m^2	33.4	37.4
More than 130 m^2	12.3	15.8
8. Building Age		
More Than 50 Years	25.5	11.3
31 – 50 Years	27.7	13.6
21 – 30 Years	11.5	25.0
11 – 20 Years	11.0	33.4
Age5 = Less Than 11 Years	24.3	16.7
9. Elevator Facilities		
None		70.3
Elevator		29.7
10. Air Conditioning		
None		97.6
Air Conditioning		2.4
11. Water Heating Fuel		
None		5.4
Wood, Coal		2.1
Gas		59.8
Electric Energy		16.2
Natural Gas		16.6
12. Heating System Fuel		
None		10.9
Wood, Coal		19.4
Electric Energy		40.6
Natural Gas		16.0
13. Cooking Fuel		
Wood, Carbon		3.4
Gas		76.8
Electric Energy		6.8

Natural Gas	13.0
14. Garden Facilities	
None	84.9
Garden	15.1
15. Building Type	
Inferior Tenement	0.3
One Floor	35.6
Two Floors	4.6
Three or more Floors	59.5
16. Swimming Pool	
None	98.8
Swimming Pool	0.2
17. Sports Area	
None	98.9
Sports Area	1.1
18. Other Community Services	
None	52.7
Some Community Services	47.3

Table B. Correlation Between the First Factor Obtained by MCA and the Physical Attributes, 1980-81 and 1990-91

Physical Attributes	1980-81	Physical Attributes	1990-91
1. Hygienic Services	0.76	1. Other Community Services	0.83
2. Water Facilities	0.75	2. Elevator	0.80
3. Telephone Facilities	0.62	3. Air Conditioned	0.80
4. Building Age	0.59	4. Sports Area	0.78
5. Heating Facilities	0.41	5. Building Type	0.73
6. Garage Facilities	0.41	6. Cooking	0.72
7. Housing Size	0.40	7. Hygienic Services	0.72
8. Electricity	0.26	8. Water Heating Combustible	0.71
		9. Water Facilities	0.70
		10. Telephone Facilities	0.69
		11. Building Age	0.68
		12. Heating Facilities	0.65
		13. Garden Facilities	0.62
		14. Garage Facilities	0.58
		15. Heating System	0.57
		16. Housing Size	0.49
		17. Swimming Pool	0.40
		18. Electricity	0.33

Table C. Ranking of Normalized Category Weights In the First Factor Obtained By MCA, 1980-81 and 1990-91

1980-81

Categories	**Weights**
No Electricity	- 10.00
No Water	- 8.41
No Hygienic Services	- 8.09
Only Cold Water	- 4.21
More Than 50 Years of Age	- 3.25
Less Than 60 m^2	- 2.59
No Telephone	- 2.17
No Heating	- 1.73
31 – 50 Years of Age	-1.49
No Garage Facilities	- 0.68
21 – 30 Years of Age	- 0.22
One Full Bathroom	- 0.07
Electricity	0.09
61 – 90 m^2	0.36
11 – 20 Years of Age	1.08
91 – 130 m^2	1.43
Cold and Hot Water	1.49
More Than 130 m^2	1.65
Telephone	2.29
Less than 11 Years of Age	2.67
Mobile, Ind. or Central Heating	2.79
Garage	3.47
More Than A Full Bathroom	4.34

1990-91

Categories	**Weights**
No Water	- 10.00
No Electricity	- 8.72
No Hygienic Services	- 8.32
No Water Heating System	- 7.16
Only Cold Water	- 6.94
Wood, Coal for Cooking	- 4.97
Inferior Tenement	- 4.91
More Than 50 Years of Age	- 3.66
No Heating	- 3.62
Less Than 60 m^2	- 3.27
No Telephone	- 3.27
One Floor Building	- 2.93
30 – 50 Years of Age	- 2.82
No Community Services	- 2.44
Wood, Coal for Heating	- 2.25
21 – 30 Years of Age	- 2.01
Two Floor Building	- 1.93
No Elevator	- 1.82
Gas For Heating	- 1.35
No Garage	- 1.02
One Bathroom	- 0.98
Gas For Water Heating	- 0.89
Elevator	- 0.81
10 – 20 Years of Age	- 0.81
Gas For Cooking	- 0.69
61 – 90 m^2	- 0.33
No Air Conditioning	- 0.08
No Sports Area	- 0.05
No Swimming Pool	- 0.04
No Garden	- 0.01
Electricity	0.01
Garden	0.04
Electric Energy for Heating	0.15
Mobile, Ind. or Central Heating	0.22
90 – 120 m^2	0.32
Cold and Hot Water	0.32
Electric Energy For Water Heating	0.41
More Than 120^2	0.81
Telephone	1.14
Less Than 11 Years Old	2.14
More Than Two Floors	2.21
Garage	2.41
Natural Gas Heating	2.65
Some Community Services	3.16
Air Conditioning	3.29
More Than One Full Bathroom	3.41
Gas For Water Heating	3.65
Electric Energy For Cooking	4.46
Swimming Pool	4.69

Elevator	4.69
Natural Gas for Cooking	5.19
Sports Area	5.49